

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: October 25, 2001, 11:43:24 : Search time 12.34 Seconds  
(without alignments)  
672,440 Million cell updates/sec

Title: US-09-412-100-23

Perfect score: 2079  
Sequence: 1 MSINTSGIASTMISTGCA.....DAMAGDAINNALIKIGAA 403

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCrus.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfilest.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2079	100.0	403	2	US-08-200-724A-2
2	2079	100.0	403	2	US-09-030-270A-3
3	2079	100.0	403	4	US-08-851-376A-2
4	2079	100.0	403	4	US-08-984-207-3
5	1928	92.7	385	1	US-08-891-254-3
6	1928	92.7	385	2	US-08-819-539-3
7	1928	92.7	385	5	PCT-US96-08819-3
8	1913	92.0	385	5	PCT-US93-06243-2
9	718.5	34.6	338	1	US-08-891-254-1
10	718.5	34.6	338	2	US-08-484-358-2
11	718.5	34.6	338	2	US-08-819-539-1
12	718.5	34.6	338	2	US-09-030-270A-1
13	718.5	34.6	338	4	US-09-118-959-2
14	718.5	34.6	338	4	US-08-984-207-1
15	718.5	34.6	338	5	PCT-US96-08819-1
16	211.5	10.2	318	4	US-09-060-756-77
17	211.5	10.2	718	1	US-08-425-069-2
18	211.5	10.2	718	2	US-08-317-844B-2
19	211.5	10.2	747	3	US-09-034-177-3
20	199.5	9.6	334	4	US-09-060-756-728
21	198.5	9.5	738	3	US-08-864-038A-3
22	197.5	9.5	344	1	US-08-891-254-7
23	197.5	9.5	344	2	US-08-819-539-7
24	197.5	9.5	344	2	US-09-030-270A-7
25	197.5	9.5	344	4	US-08-984-207-7
26	197.5	9.5	344	5	PCT-US96-08819-7
27	194	9.3	1160	3	US-08-808-559A-24

28	186	8.9	749	1	US-08-317-522A-2	Sequence 2, Appli
29	186	8.9	749	1	US-08-439-818A-2	Sequence 2, Appli
30	186	8.9	749	2	US-08-751-965-2	Sequence 2, Appli
31	186	8.9	749	2	US-08-738-975-2	Sequence 2, Appli
32	186	8.9	749	2	US-08-728-626-2	Sequence 2, Appli
33	186	8.9	749	3	US-08-808-559A-2	Sequence 2, Appli
34	185.5	8.8	235	2	US-08-529-108B-1	Sequence 1, Appli
35	183	8.8	674	1	US-08-317-522A-3	Sequence 3, Appli
36	183	8.8	674	1	US-08-439-818A-3	Sequence 3, Appli
37	183	8.8	674	2	US-08-751-965-3	Sequence 3, Appli
38	183	8.8	674	2	US-08-738-975-3	Sequence 3, Appli
39	183	8.8	674	2	US-08-728-626-3	Sequence 3, Appli
40	183	8.8	674	3	US-08-808-559A-3	Sequence 3, Appli
41	171.5	8.2	745	2	US-09-010-928B-28	Sequence 28, Appli
42	171.5	8.2	870	2	US-09-010-928B-4	Sequence 2, Appli
43	171.5	8.2	907	2	US-09-010-928B-4	Sequence 4, Appli
44	171	8.2	1177	1	US-07-609-716-31	Sequence 31, Appli
45	171	8.2	1177	1	US-08-175-155-29	Sequence 29, Appli

## ALIGNMENTS

RESULT 1  
US-08-200-724A-2  
Sequence 2, Application US/08200724A  
Patent No. 5849868  
GENERAL INFORMATION:  
APPLICANT: Wei, Zhong-Min  
APPLICANT: Bauer, David W.  
APPLICANT: Beer, Steven V.  
APPLICANT: Collier, Alan  
APPLICANT: He, Sheng-Yang  
APPLICANT: Laby, Ron J.  
TITLE OF INVENTION: ELICITOR OF THE HYPERSENSITIVE RESPONSE  
TITLE OF INVENTION: IN PLANTS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
STREET: Clinton Square  
CITY: Rochester  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/200.724A  
FILING DATE: 23-FEB-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldman, Michael L.  
REGISTRATION NUMBER: 30,727  
REFERENCE/DOCKET NUMBER: 19603/10030  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1304  
TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 403 amino acids  
TYPE: amino acid  
STRANDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-200-724A-2  
Query Match 100.0% Score 2079 DB 2 Length 403:  
Best Local Similarity 100.0% Pred. No. 5.4e-175:  
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MSINTSGLASTMQISIGAGAGNNGLGSTRONAGLGSNALSIGGSGNDVTYNOLAGL 60
Db 1 MSINTSGLASTMQISIGAGAGNNGLGSTRONAGLGSNALSIGGSGNDVTYNOLAGL 60
QY 61 TGMAMMAMMAGGGLMGCGGLGAGLGNGLGSGGGLGEBLSNALNDMLGSLNTLGSKGGN 120
Db 61 TGMAMMAMMAGGGLMGCGGLGAGLGNGLGSGGGLGEBLSNALNDMLGSLNTLGSKGGN 120
QY 121 TTTSTNSPLDQALGINSTSONDSTSGTSDSPMQLLMFSEIMOSLFGDGDGT 180
Db 121 TTTSTNSPLDQALGINSTSONDSTSGTSDSPMQLLMFSEIMOSLFGDGDGT 180
QY 181 QGSSSGKOPTBEBQNAKKGVTDALSLGMLNGLSQLLGNGLGSGGAGAGTGLDSSSL 240
Db 181 QGSSSGKOPTBEBQNAKKGVTDALSLGMLNGLSQLLGNGLGSGGAGAGTGLDSSSL 240
QY 241 GKGGLQMLSGPVYQOLGNAVGTGIGMKAGIOALNDIGTRHSHSTRSFVKKGRAMAKEI 300
Db 241 GKGGLQMLSGPVYQOLGNAVGTGIGMKAGIOALNDIGTRHSHSTRSFVKKGRAMAKEI 300
QY 301 GGFMDQPEVFGKPOYOKGPGOEKTDKSWAKALSKPDDGMPASMEQFNRAKGMIR 360
Db 301 GGFMDQPEVFGKPOYOKGPGOEKTDKSWAKALSKPDDGMPASMEQFNRAKGMIR 360
QY 361 PMAGDTGNGMLQARGAGSSSLGIDAMMAGDAINNMALGKLGAA 403
Db 361 PMAGDTGNGMLQARGAGSSSLGIDAMMAGDAINNMALGKLGAA 403

```

```

RESULT 2
US-09-030-270A-3
; Sequence 3, Application us/09030270A
; Patent No. 5977060
; GENERAL INFORMATION:
; APPLICANT: Zitter, Thomas A.
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: INSECT CONTROL WITH A
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Haigraue, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,270A
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/039,226
; FILING DATE: 28-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1521
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

```

US-09-030-270A-3
Query Match 100.0%; Score 2079; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 5,4e-175;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSINTSGLASTMQISIGAGAGNNGLGSTRONAGLGSNALSIGGSGNDVTYNOLAGL 60
Db 1 MSINTSGLASTMQISIGAGAGNNGLGSTRONAGLGSNALSIGGSGNDVTYNOLAGL 60
QY 61 TGMAMMAMMAGGGLMGCGGLGAGLGNGLGSGGGLGEBLSNALNDMLGSLNTLGSKGGN 120
Db 61 TGMAMMAMMAGGGLMGCGGLGAGLGNGLGSGGGLGEBLSNALNDMLGSLNTLGSKGGN 120
QY 121 TTTSTNSPLDQALGINSTSONDSTSGTSDSPMQLLMFSEIMOSLFGDGDGT 180
Db 121 TTTSTNSPLDQALGINSTSONDSTSGTSDSPMQLLMFSEIMOSLFGDGDGT 180
QY 181 QGSSSGKOPTBEBQNAKKGVTDALSLGMLNGLSQLLGNGLGSGGAGAGTGLDSSSL 240
Db 181 QGSSSGKOPTBEBQNAKKGVTDALSLGMLNGLSQLLGNGLGSGGAGAGTGLDSSSL 240
QY 241 GKGGLQMLSGPVYQOLGNAVGTGIGMKAGIOALNDIGTRHSHSTRSFVKKGRAMAKEI 300
Db 241 GKGGLQMLSGPVYQOLGNAVGTGIGMKAGIOALNDIGTRHSHSTRSFVKKGRAMAKEI 300
QY 301 GGFMDQPEVFGKPOYOKGPGOEKTDKSWAKALSKPDDGMPASMEQFNRAKGMIR 360
Db 301 GGFMDQPEVFGKPOYOKGPGOEKTDKSWAKALSKPDDGMPASMEQFNRAKGMIR 360
QY 361 PMAGDTGNGMLQARGAGSSSLGIDAMMAGDAINNMALGKLGAA 403
Db 361 PMAGDTGNGMLQARGAGSSSLGIDAMMAGDAINNMALGKLGAA 403

```

```

RESULT 3
US-08-851-376A-2
; Sequence 2, Application US/08851376A
; Patent No. 6174717
; GENERAL INFORMATION:
; APPLICANT: Beer, Steven V.
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Bauer, David W.
; APPLICANT: Collmer, Alan
; APPLICANT: He, Sheng-Yang
; APPLICANT: Lady, Ron
; TITLE OF INVENTION: ELICITOR OF THE HYPERSENSITIVE RESPONSE
; TITLE OF INVENTION: IN PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon Peabody LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851,376A
; FILING DATE: 05-MAY-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/200,724
; FILING DATE: 23-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/10035

```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1304  
TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 403 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-851-376A-2

Query Match 100.0%; Score 2079; DB 4; Length 403;  
Best Local Similarity 100.0%; Pred. No. 5.4e-175;  
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLNTSGIGASTMIOISIGAGGNNGLTSTRONAGLGSNSALGAGGNDVTNQLAGLL 60  
DB 1 MSLNTSGIGASTMIOISIGAGGNNGLTSTRONAGLGSNSALGAGGNDVTNQLAGLL 60  
QY 61 TGMAMMMSMMGGGGLMGGLGGGLGNGLSGSGGLEGLSNALNDMLGSLNTLGSKGGNN 120  
DB 61 TGMAMMMSMMGGGGLMGGLGGGLGNGLSGSGGLEGLSNALNDMLGSLNTLGSKGGNN 120  
QY 121 TTSTTNSPLDQALGINTSQNDSTSGTDTSDSSDPMQQLKMFSEIMOSLFGDGQDGT 180  
DB 121 TTSTTNSPLDQALGINTSQNDSTSGTDTSDSSDPMQQLKMFSEIMOSLFGDGQDGT 180  
QY 181 OGSSSGGKOPTGEGEONAKKVTDALSGLMNGLSQLLNGGLGGGCGNAGTGLDGSSL 240  
DB 181 OGSSSGGKOPTGEGEONAKKVTDALSGLMNGLSQLLNGGLGGGCGNAGTGLDGSSL 240  
QY 241 GKGGLQNLGSPVYQQLGNAVGTGIGMKAGIQALNDIGTHRRSSTRSVNKGDRAMAEI 300  
DB 241 GKGGLQNLGSPVYQQLGNAVGTGIGMKAGIQALNDIGTHRRSSTRSVNKGDRAMAEI 300  
QY 301 GQFMDOYFEVFGKPOYQKPGQEVKTDKSWAKALSKPDDGMPASMEQFNKAKGMIR 360  
DB 301 GQFMDOYFEVFGKPOYQKPGQEVKTDKSWAKALSKPDDGMPASMEQFNKAKGMIR 360  
QY 361 PMAGDTGNLQARGAGSSSLGIDAMAGDAINNALGKLGAA 403  
DB 361 PMAGDTGNLQARGAGSSSLGIDAMAGDAINNALGKLGAA 403

## RESULT 4

US-08-984-207-3  
Sequence 3, Application US/08984207  
Patent No. 6235974

GENERAL INFORMATION:  
APPLICANT: Qiu, Dewen  
APPLICANT: Wei, Zhong-Min  
APPLICANT: Beer, Steven V  
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED  
TITLE OF INVENTION: RESISTANCE IN PLANTS BY SEED TREATMENT  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
STREET: P.O. Box 1051, Clinton Square  
CITY: Rochester  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/984,207  
FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/033,230  
FILING DATE: 05-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldman, Michael L.  
REGISTRATION NUMBER: 30,727  
REFERENCE/DOCKET NUMBER: 19603/1201  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1304  
TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 403 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-984-207-3

Query Match 100.0%; Score 2079; DB 4; Length 403;  
Best Local Similarity 100.0%; Pred. No. 5.4e-175;  
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLNTSGIGASTMIOISIGAGGNNGLTSTRONAGLGSNSALGAGGNDVTNQLAGLL 60  
DB 1 MSLNTSGIGASTMIOISIGAGGNNGLTSTRONAGLGSNSALGAGGNDVTNQLAGLL 60  
QY 61 TGMAMMMSMMGGGGLMGGLGGGLGNGLSGSGGLEGLSNALNDMLGSLNTLGSKGGNN 120  
DB 61 TGMAMMMSMMGGGGLMGGLGGGLGNGLSGSGGLEGLSNALNDMLGSLNTLGSKGGNN 120  
QY 121 TTSTTNSPLDQALGINTSQNDSTSGTDTSDSSDPMQQLKMFSEIMOSLFGDGQDGT 180  
DB 121 TTSTTNSPLDQALGINTSQNDSTSGTDTSDSSDPMQQLKMFSEIMOSLFGDGQDGT 180  
QY 181 OGSSSGGKOPTGEGEONAKKVTDALSGLMNGLSQLLNGGLGGGCGNAGTGLDGSSL 240  
DB 181 OGSSSGGKOPTGEGEONAKKVTDALSGLMNGLSQLLNGGLGGGCGNAGTGLDGSSL 240  
QY 241 GKGGLQNLGSPVYQQLGNAVGTGIGMKAGIQALNDIGTHRRSSTRSVNKGDRAMAEI 300  
DB 241 GKGGLQNLGSPVYQQLGNAVGTGIGMKAGIQALNDIGTHRRSSTRSVNKGDRAMAEI 300  
QY 301 GQFMDOYFEVFGKPOYQKPGQEVKTDKSWAKALSKPDDGMPASMEQFNKAKGMIR 360  
DB 301 GQFMDOYFEVFGKPOYQKPGQEVKTDKSWAKALSKPDDGMPASMEQFNKAKGMIR 360  
QY 361 PMAGDTGNLQARGAGSSSLGIDAMAGDAINNALGKLGAA 403  
DB 361 PMAGDTGNLQARGAGSSSLGIDAMAGDAINNALGKLGAA 403

## RESULT 5

US-08-891-254-3  
Sequence 3, Application US/08891254  
Patent No. 5776889

GENERAL INFORMATION:  
APPLICANT: Wei, Zhong-Min  
APPLICANT: Beer, Steven V  
TITLE OF INVENTION: Hypersensitive Response  
TITLE OF INVENTION: Induced Resistance in Plants  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
STREET: Clinton Square, P.O. Box 1051  
CITY: Rochester  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/891,254  
FILING DATE: 10-JUL-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/475,775  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldman, Michael L.  
REGISTRATION NUMBER: 30,727  
REFERENCE/DOCKET NUMBER: 14603/10050  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1304  
TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 385 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-891-254-3

Query Match 92.7%; Score 1928; DB 1; Length 385;  
Best Local Similarity 100.0%; Pred. No. 9.4e-162;  
Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1<sup>1</sup> MSLNTSGLAGSTWQISIGAGAGNNGLLGTSRONAGLGNSALGAGGNQNDTVNQLAGLL 60  
DB 1 MSLNTSGLAGSTWQISIGAGAGNNGLLGTSRONAGLGNSALGAGGNQNDTVNQLAGLL 60  
QY 61 TGMAMMMMSMGGGGLMGGLGAGGGLGNGLGSGGGLGEGLSNALNDMLGSLNTLTGSKGNN 120  
DB 61 TGMAMMMMSMGGGGLMGGLGAGGGLGNGLGSGGGLGEGLSNALNDMLGSLNTLTGSKGNN 120  
QY 121 TTTSTNSPLDQALGINSQNDSTSGTSDSPDPMOQLLKMFSITMOSLFGDGDGT 180  
DB 121 TTTSTNSPLDQALGINSQNDSTSGTSDSPDPMOQLLKMFSITMOSLFGDGDGT 180  
QY 121 TTTSTNSPLDQALGINSQNDSTSGTSDSPDPMOQLLKMFSITMOSLFGDGDGT 180  
DB 121 TTTSTNSPLDQALGINSQNDSTSGTSDSPDPMOQLLKMFSITMOSLFGDGDGT 180  
QY 181 QGSSSGGKQPTGEGBNAYKKGVTDALSGLMGNGLSQLLNGGGLGAGGAGNAGTGLDSSL 240  
DB 181 QGSSSGGKQPTGEGBNAYKKGVTDALSGLMGNGLSQLLNGGGLGAGGAGNAGTGLDSSL 240  
QY 241 GGGKGLQNLSCPVYQOJGNAVGTGIGMKAGIQALNDIGTRHSSSTRSFVNGDRAMAKEY 300  
DB 241 GGGKGLQNLSCPVYQOJGNAVGTGIGMKAGIQALNDIGTRHSSSTRSFVNGDRAMAKEY 300  
QY 301 GGFMDQYPEVFGKPOYKPGQEVKTDKSWAKALSKPDDGGMTPASMEOPNKAKEY 360  
DB 301 GGFMDQYPEVFGKPOYKPGQEVKTDKSWAKALSKPDDGGMTPASMEOPNKAKEY 360  
QY 361 PMAGDTGNGNLQ 372  
DB 361 PMAGDTGNGNLQ 372

RESULT 6  
US-08-819-539-3  
Sequence 3, Application US/08819539  
Patent No. 5859324  
GENERAL INFORMATION:  
APPLICANT: Wei, Zhong-Min  
APPLICANT: Beer, Steven V.  
TITLE OF INVENTION: Hypersensitive Response  
TITLE OF INVENTION: Induced Resistance in Plants  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
STREET: Clinton Square, P.O. Box 1051  
CITY: Rochester

STATE: New York  
COUNTRY: U.S.A.  
ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/819,539  
FILING DATE: 17-MAR-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/475,775  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldman, Michael L.  
REGISTRATION NUMBER: 30,727  
REFERENCE/DOCKET NUMBER: 14603/10050  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1304  
TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 385 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-819-539-3

Query Match 92.7%; Score 1928; DB 2; Length 385;  
Best Local Similarity 100.0%; Pred. No. 9.4e-162;  
Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1<sup>1</sup> MSLNTSGLAGSTWQISIGAGAGNNGLLGTSRONAGLGNSALGAGGNQNDTVNQLAGLL 60  
DB 1 MSLNTSGLAGSTWQISIGAGAGNNGLLGTSRONAGLGNSALGAGGNQNDTVNQLAGLL 60  
QY 61 TGMAMMMMSMGGGGLMGGLGAGGGLGNGLGSGGGLGEGLSNALNDMLGSLNTLTGSKGNN 120  
DB 61 TGMAMMMMSMGGGGLMGGLGAGGGLGNGLGSGGGLGEGLSNALNDMLGSLNTLTGSKGNN 120  
QY 121 TTTSTNSPLDQALGINSQNDSTSGTSDSPDPMOQLLKMFSITMOSLFGDGDGT 180  
DB 121 TTTSTNSPLDQALGINSQNDSTSGTSDSPDPMOQLLKMFSITMOSLFGDGDGT 180  
QY 121 TTTSTNSPLDQALGINSQNDSTSGTSDSPDPMOQLLKMFSITMOSLFGDGDGT 180  
DB 121 TTTSTNSPLDQALGINSQNDSTSGTSDSPDPMOQLLKMFSITMOSLFGDGDGT 180  
QY 181 QGSSSGGKQPTGEGBNAYKKGVTDALSGLMGNGLSQLLNGGGLGAGGAGNAGTGLDSSL 240  
DB 181 QGSSSGGKQPTGEGBNAYKKGVTDALSGLMGNGLSQLLNGGGLGAGGAGNAGTGLDSSL 240  
QY 241 GGGKGLQNLSCPVYQOJGNAVGTGIGMKAGIQALNDIGTRHSSSTRSFVNGDRAMAKEY 300  
DB 241 GGGKGLQNLSCPVYQOJGNAVGTGIGMKAGIQALNDIGTRHSSSTRSFVNGDRAMAKEY 300  
QY 301 GGFMDQYPEVFGKPOYKPGQEVKTDKSWAKALSKPDDGGMTPASMEOPNKAKEY 360  
DB 301 GGFMDQYPEVFGKPOYKPGQEVKTDKSWAKALSKPDDGGMTPASMEOPNKAKEY 360  
QY 361 PMAGDTGNGNLQ 372  
DB 361 PMAGDTGNGNLQ 372

RESULT 7  
PCT-US96-08819-3  
Sequence 3, Application PC/TUS9608819  
GENERAL INFORMATION:  
APPLICANT: Cornell Research Foundation, Inc.  
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED  
TITLE OF INVENTION: RESISTANCE IN PLANTS  
NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
 STREET: Clinton Square, P.O. Box 1051  
 CITY: Rochester  
 STATE: New York  
 COUNTRY: U.S.A.  
 ZIP: 14603  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentln Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US96/08819  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/475,775  
 FILING DATE: 07-JUN-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Goldman, Michael L.  
 REGISTRATION NUMBER: 30,727  
 REFERENCE/DOCKET NUMBER: 19603/10051  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (716) 263-1304  
 TELEFAX: (716) 263-1600  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 385 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 PCT-US96-08819-3

Query Match 92.7%; Score 1928; DB 5; Length 385;  
 Best Local Similarity 100.0%; Pred. No. 9,4e-162;  
 Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLSNGLGASTMOJISIGAGGNNGLGTSRONAGIGNSALGIGGNNONTVNOAGLL 60  
 DB 1 MSLSNGLGASTMOJISIGAGGNNGLGTSRONAGIGNSALGIGGNNONTVNOAGLL 60  
 QY 61 TGMNMMMSMGGGGLMGGLGGGLGNGLGSGGLGEGLSNALNDMLGSLNTLGSKGGNN 120  
 DB 61 TGMNMMMSMGGGGLMGGLGGGLGNGLGSGGLGEGLSNALNDMLGSLNTLGSKGGNN 120  
 QY 121 TTSTNSPLDQALGINSTSQNDSTSGTDTSDSSDPMOQLKMFSEIMQSLFGGQDGT 180  
 DB 121 TTSTNSPLDQALGINSTSQNDSTSGTDTSDSSDPMOQLKMFSEIMQSLFGGQDGT 180  
 QY 181 OGSSSGGKOPTGEBONAYKKVTDALSGLMNGLSQLLNGGLGGGOGGNACTGLDSSL 240  
 DB 181 OGSSSGGKOPTGEBONAYKKVTDALSGLMNGLSQLLNGGLGGGOGGNACTGLDSSL 240  
 QY 241 GKGGLQNLGSPYDQOLGNAVGTGIGMKAGIOALNDIGTHRSSTFSFVNKGDRAKAKET 300  
 DB 241 GKGGLQNLGSPYDQOLGNAVGTGIGMKAGIOALNDIGTHRSSTFSFVNKGDRAKAKET 300  
 QY 301 GQFMDQYEVFEFGKPOYOKGPGQEVKTDKSNKALSKPDDGMPASMOQFNKAKGMIR 360  
 DB 301 GQFMDQYEVFEFGKPOYOKGPGQEVKTDKSNKALSKPDDGMPASMOQFNKAKGMIR 360  
 QY 361 PMAGDTGNGNL 372  
 DB 361 PMAGDTGNGNL 372

RESULT 8  
 PCT-US93-06243-2  
 Sequence 2, Application PC/TUS9306243  
 GENERAL INFORMATION:

APPLICANT: Zhong-Min Wei, David M. Bauer, Steven V.  
 APPLICANT: Beer, Alan Collier, Sheng-Yang He, and Ron J. Laby  
 TITLE OF INVENTION: Elicitor of the Hypersensitive Response in Plants  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Yahwak & Associates  
 STREET: 25 Skytop Drive  
 CITY: Trumbull  
 STATE: Connecticut  
 COUNTRY: USA  
 ZIP: 06611  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: Macintosh  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: Microsoft Word 4.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US93/06243  
 FILING DATE: 19930630  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 907,935  
 FILING DATE: 01-JUL-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: George M. Yahwak  
 REGISTRATION NUMBER: 26,824  
 REFERENCE/DOCKET NUMBER: CRF D-1172  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (203)268-1951  
 TELEFAX: (203)268-1951  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 385 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 PCT-US93-06243-2

Query Match 92.0%; Score 1913; DB 5; Length 385;  
 Best Local Similarity 99.5%; Pred. No. 1.9e-160;  
 Matches 369; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSLSNGLGASTMOJISIGAGGNNGLGTSRONAGIGNSALGIGGNNONTVNOAGLL 60  
 DB 1 MSLSNGLGASTMOJISIGAGGNNGLGTSRONAGIGNSALGIGGNNONTVNOAGLL 60  
 QY 61 TGMNMMMSMGGGGLMGGLGGGLGNGLGSGGLGEGLSNALNDMLGSLNTLGSKGGNN 120  
 DB 61 TGMNMMMSMGGGGLMGGLGGGLGNGLGSGGLGEGLSNALNDMLGSLNTLGSKGGNN 120  
 QY 121 TTSTNSPLDQALGINSTSQNDSTSGTDTSDSSDPMOQLKMFSEIMQSLFGGQDGT 180  
 DB 121 TTSTNSPLDQALGINSTSQNDSTSGTDTSDSSDPMOQLKMFSEIMQSLFGGQDGT 180  
 QY 181 OGSSSGGKOPTGEBONAYKKVTDALSGLMNGLSQLLNGGLGGGOGGNACTGLDSSL 240  
 DB 181 OGSSSGGKOPTGEBONAYKKVTDALSGLMNGLSQLLNGGLGGGOGGNACTGLDSSL 240  
 QY 241 GKGGLQNLGSPYDQOLGNAVGTGIGMKAGIOALNDIGTHRSSTFSFVNKGDRAKAKET 300  
 DB 241 GKGGLQNLGSPYDQOLGNAVGTGIGMKAGIOALNDIGTHRSSTFSFVNKGDRAKAKET 300  
 QY 301 GQFMDQYEVFEFGKPOYOKGPGQEVKTDKSNKALSKPDDGMPASMOQFNKAKGMIR 360  
 DB 301 GQFMDQYEVFEFGKPOYOKGPGQEVKTDKSNKALSKPDDGMPASMOQFNKAKGMIR 360  
 QY 361 PMAGDTGNGNL 371  
 DB 361 PMAGDTGNGNL 371



QY 300 IGFMDQYPEYFGKPOYOKGQEVKTDKSWAKALSKPDDGKTPASMEOPNKAQK 359  
Db 235 IGFMDQYPEYFGKPOYOKGQEVKTDKSWAKALSKPDDGKTPASMEOPNKAQK 294  
QY 360 RPMAGDTGNTNLARGAGSSLGIDAMAGDAINMALGKLGAA 403  
Db 295 SAVAGDTGNTNLARGAGSSLGIDAMAGDAINMALGKLGAA 338

## RESULT 11

US-08-819-539-1  
Sequence 1, Application US/08819539  
Patent No. 5859324  
GENERAL INFORMATION:  
APPLICANT: Wei, Zhong-Min  
APPLICANT: Beer, Steven V.  
TITLE OF INVENTION: Hypersensitive Response  
TITLE OF INVENTION: Induced Resistance in Plants  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
STREET: Clinton Square, P.O. Box 1051  
CITY: Rochester  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/819,539  
FILING DATE: 17-MAR-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/475,775  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldman, Michael L.  
REGISTRATION NUMBER: 30,727  
REFERENCE/DOCKET NUMBER: 14603/10050  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1304  
TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 338 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-819-539-1

Query Match 34.6%; Score 718.5; DB 2; Length 338;  
Best Local Similarity 42.8%; Pred. No. 1.4e-55;  
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;

QY 13 MOIST-GGAGGNNGLGTSRONAGIG-NSA---LGLGGGNDNTVNLGLLGTGMNMM 67  
Db 1 MOITKAHIGDGLGVSGAGQ--GLKGLNSAASSLGSSVDKLSITDKLTSALTSMM--- 55  
QY 68 SMAGGGGLMGGGLGGLGNGLGSGGLGEGLSNALNDMLGSLNLTGSKGNNNTSTNS 127  
Db 56 -----FGALAOGLGAS--SKGIGMSNOLGOSFGN-----GAQASNLISVPR- 96  
QY 128 PLDQALINSTSNDSTSGTSDTSDSDPMQOLKMFSEIMQSLFG-----DGQDG 179  
Db 97 -----SGQDALS-----KMFQKALDDLLGHDTYTKLNQSNQ 128  
QY 180 TQSSSGGKQPTGGRONAYKKGVTDALSLGMLGNGLSQLLGNGLGGGCGNAGTGLDSS 239

Db 129 LANSMLNASQMTQGNMNAFSGGVNNAALSLTNGNLGQSM-----SGFSOPS 174  
QY 240 LGGKGLNLSPGVYOOGLNVAVTGICMKACIQALNDIGTRHSSTSFVNKGRANAKE 299  
Db 175 LGAGLGGLSGAGAFNOLGNALIGMGVGNALSLNSVSTHVDGNNRHFVDKEDRGNAKE 234  
QY 300 IGFMDQYPEYFGKPOYOKGQEVKTDKSWAKALSKPDDGKTPASMEOPNKAQK 359  
Db 235 IGFMDQYPEYFGKPOYOKGQEVKTDKSWAKALSKPDDGKTPASMEOPNKAQK 294  
QY 360 RPMAGDTGNTNLARGAGSSLGIDAMAGDAINMALGKLGAA 403  
Db 295 SAVAGDTGNTNLARGAGSSLGIDAMAGDAINMALGKLGAA 338

## RESULT 12

US-09-030-270A-1  
Sequence 1, Application US/09030270A  
Patent No. 5977060  
GENERAL INFORMATION:  
APPLICANT: Zitter, Thomas A.  
APPLICANT: Wei, Zhong-Min  
TITLE OF INVENTION: INSECT CONTROL WITH A  
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
STREET: P.O. Box 1051, Clinton Square  
CITY: Rochester  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/030,270A  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/039,226  
FILING DATE: 28-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldman, Michael L.  
REGISTRATION NUMBER: 30,727  
REFERENCE/DOCKET NUMBER: 19603/1521  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1304  
TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 338 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-030-270A-1

Query Match 34.6%; Score 718.5; DB 2; Length 338;  
Best Local Similarity 42.8%; Pred. No. 1.4e-55;  
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;

QY 13 MOIST-GGAGGNNGLGTSRONAGIG-NSA---LGLGGGNDNTVNLGLLGTGMNMM 67  
Db 1 MOITKAHIGDGLGVSGAGQ--GLKGLNSAASSLGSSVDKLSITDKLTSALTSMM--- 55  
QY 68 SMAGGGGLMGGGLGGLGNGLGSGGLGEGLSNALNDMLGSLNLTGSKGNNNTSTNS 127  
Db 56 -----FGALAOGLGAS--SKGIGMSNOLGOSFGN-----GAQASNLISVPR- 96

QY 128 PLDQALGINSTSDSTGSDSSDPMOQLKMFSEIMOSLFG-----DGQDG 179  
DB 97 -----SGGDALS-----KMFDRKALDDLHHDVTKLTLNQSNO 128  
QY 180 TQSSSGGKOPTBEONAYKKGYVDALSGLMGNGLSOLLGNGGAGGOGGAGTGLDSS 239  
DB 129 LANSMLNASQMTQGNMNAFSGGVNNAISLILGNGLGSM-----SGFSQPS 174  
QY 240 LGGKGLQNLGSPVDYQOOLGNAVGTGIGMKAGIQALNDIGTHRHSTSFVYKGRAPAKE 299  
DB 175 LGAGGLQGLSAGAFNQLGNAIGMGVGNALSLSVSTHVDSNNHFFVDEKEDRGMAKE 234  
QY 300 IGFPMDOYPEYFGKPYOKGPGQEVKTDKSWAKALSKPDDGWTSPASMEQFNKAKGMK 359  
DB 235 IGFPMDOYPEYFGKPYOKGPGQEVKTDKSWAKALSKPDDGWTSPASMEQFNKAKGMK 294  
QY 360 RPAAGDTGNGNLQARGAGSSIGIDAMAGDAINNALGKIGAA 403  
DB 295 SAVAGDTGNTNLNRGAGASLGIDAAYGDKIANMSLIGKILANA 338  
RESULT 13  
US-09-118-959-2  
; Sequence 2, Application US/09118959  
; Patent No. 6001959  
; GENERAL INFORMATION:  
; APPLICANT: Bauer, David  
; APPLICANT: Collmer, Alan  
; TITLE OF INVENTION: Hypersensitive Response Elicitor From  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
; STREET: Clinton Square  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/118,959  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldmen, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/840  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 716-263-1304  
; TELEFAX: 716-263-1600  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 338 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-118-959-2  
Query Match 34.6%; Score 718.5; DB 3; Length 338;  
Best Local Similarity 42.8%; Pred. No. 1,4e-55;  
Matches 1/3; Conservative 41; Mismatches 11; Indels 79; Gaps 11;  
QY 13 MOISI-GGAGCNGCLGTSRQNGGLG-NSA---LGAGGQNGQNTVQNLGGLTGNMAM 67  
DB 1 MOTTIRAHIGDLGVSLGAG--GLKGLNSPAASSLGSSVDKLSITDKTSALTSMA--- 55

QY 68 SMGCGGIMGGGLGGLGNGLGSGGLGEGLSNALNDMLTGSNTLTGSKGNNSTSTNS 127  
DB 56 -----FGALAQIGIGAS-SKGLGMSNOLGSGFGN-----GAQASNLISLPK- 96  
QY 128 PLDQALGINSTSDSTGSDSSDPMOQLKMFSEIMOSLFG-----DGQDG 179  
DB 97 -----SGGDALS-----KMFDRKALDDLHHDVTKLTLNQSNO 128  
QY 180 TQSSSGGKOPTBEONAYKKGYVDALSGLMGNGLSOLLGNGGAGGOGGAGTGLDSS 239  
DB 129 LANSMLNASQMTQGNMNAFSGGVNNAISLILGNGLGSM-----SGFSQPS 174  
QY 240 LGGKGLQNLGSPVDYQOOLGNAVGTGIGMKAGIQALNDIGTHRHSTSFVYKGRAPAKE 299  
DB 175 LGAGGLQGLSAGAFNQLGNAIGMGVGNALSLSVSTHVDSNNHFFVDEKEDRGMAKE 234  
QY 300 IGFPMDOYPEYFGKPYOKGPGQEVKTDKSWAKALSKPDDGWTSPASMEQFNKAKGMK 359  
DB 235 IGFPMDOYPEYFGKPYOKGPGQEVKTDKSWAKALSKPDDGWTSPASMEQFNKAKGMK 294  
QY 360 RPAAGDTGNGNLQARGAGSSIGIDAMAGDAINNALGKIGAA 403  
DB 295 SAVAGDTGNTNLNRGAGASLGIDAAYGDKIANMSLIGKILANA 338  
RESULT 14  
US-08-984-207-1  
; Sequence 1, Application US/08984207  
; Patent No. 623574  
; GENERAL INFORMATION:  
; APPLICANT: Qiu, Dewen  
; APPLICANT: Wei, Zhong-Min  
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: P.O. Box 1051, Clinton Square  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/984,207  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/033,230  
; FILING DATE: 05-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/1201  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (716) 263-1304  
; TELEFAX: (716) 263-1600  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 338 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-984-207-1  
Query Match 34.6%; Score 718.5; DB 4; Length 338;





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 11:43:24 ; Search time 22.79 Seconds

(without alignments)  
1072.026 Million cell updates/sec

Title: US-09-412-100-23

Perfect score: 2079  
Sequence: 1 MSINISGLASTWQISIGGA.....DAMAGDAINNALGLGAA 403

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: /SID58/gcgdata/geneseq/geneseqp/AA1980.DAT:\*  
2: /SID58/gcgdata/geneseq/geneseqp/AA1981.DAT:\*  
3: /SID58/gcgdata/geneseq/geneseqp/AA1982.DAT:\*  
4: /SID58/gcgdata/geneseq/geneseqp/AA1983.DAT:\*  
5: /SID58/gcgdata/geneseq/geneseqp/AA1984.DAT:\*  
6: /SID58/gcgdata/geneseq/geneseqp/AA1985.DAT:\*  
7: /SID58/gcgdata/geneseq/geneseqp/AA1986.DAT:\*  
8: /SID58/gcgdata/geneseq/geneseqp/AA1987.DAT:\*  
9: /SID58/gcgdata/geneseq/geneseqp/AA1988.DAT:\*  
10: /SID58/gcgdata/geneseq/geneseqp/AA1989.DAT:\*  
11: /SID58/gcgdata/geneseq/geneseqp/AA1990.DAT:\*  
12: /SID58/gcgdata/geneseq/geneseqp/AA1991.DAT:\*  
13: /SID58/gcgdata/geneseq/geneseqp/AA1992.DAT:\*  
14: /SID58/gcgdata/geneseq/geneseqp/AA1993.DAT:\*  
15: /SID58/gcgdata/geneseq/geneseqp/AA1994.DAT:\*  
16: /SID58/gcgdata/geneseq/geneseqp/AA1995.DAT:\*  
17: /SID58/gcgdata/geneseq/geneseqp/AA1996.DAT:\*  
18: /SID58/gcgdata/geneseq/geneseqp/AA1997.DAT:\*  
19: /SID58/gcgdata/geneseq/geneseqp/AA1998.DAT:\*  
20: /SID58/gcgdata/geneseq/geneseqp/AA1999.DAT:\*  
21: /SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:\*  
22: /SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2079	100.0	403	AAW75863	Erwinia amylovora
2	2079	100.0	403	AAW62455	Erwinia amylovora
3	2079	100.0	403	AAW61114	Hypersensitive res
4	2079	100.0	403	AAW87639	A hypersensitive r
5	2079	100.0	403	AAW71093	Erwinia amylovora
6	2079	100.0	403	AAV84854	A hypersensitive r
7	1928	92.7	385	AAW06598	Hypersensitive res
8	1913	92.0	385	AAW45751	Erwinia amylovora
9	718.5	34.6	338	AAW06597	Hypersensitive res
10	718.5	34.6	338	AAW87638	A hypersensitive r
11	718.5	34.6	338	AAW82407	E. chrysanthemi Hr

12	718.5	34.6	338	21	AAW71092
13	718.5	34.6	338	21	AAW84853
14	718.5	34.6	338	21	AAW55801
15	718.5	34.6	340	19	AAW75862
16	718.5	34.6	340	19	AAW61113
17	718.5	34.3	340	19	AAW62454
18	718.5	34.3	340	19	AAW31853
19	718.5	34.3	340	19	AAW31853
20	718.5	34.3	340	19	AAW31853
21	718.5	34.3	340	19	AAW31853
22	718.5	34.3	340	19	AAW31853
23	718.5	34.3	340	19	AAW31853
24	718.5	34.3	340	19	AAW31853
25	718.5	34.3	340	19	AAW31853
26	718.5	34.3	340	19	AAW31853
27	718.5	34.3	340	19	AAW31853
28	718.5	34.3	340	19	AAW31853
29	718.5	34.3	340	19	AAW31853
30	718.5	34.3	340	19	AAW31853
31	718.5	34.3	340	19	AAW31853
32	718.5	34.3	340	19	AAW31853
33	718.5	34.3	340	19	AAW31853
34	718.5	34.3	340	19	AAW31853
35	718.5	34.3	340	19	AAW31853
36	718.5	34.3	340	19	AAW31853
37	718.5	34.3	340	19	AAW31853
38	718.5	34.3	340	19	AAW31853
39	718.5	34.3	340	19	AAW31853
40	718.5	34.3	340	19	AAW31853
41	718.5	34.3	340	19	AAW31853
42	718.5	34.3	340	19	AAW31853
43	718.5	34.3	340	19	AAW31853
44	718.5	34.3	340	19	AAW31853
45	718.5	34.3	340	19	AAW31853

#### ALIGNMENTS

##### RESULT 1

AAW75863	Standard; Protein; 403 AA.
ID	AAW75863
AC	AAW75863
XX	
DT	07-DEC-1998 (first entry)
XX	
DE	Erwinia amylovora hypersensitive response elicitor (HRE).
XX	
KW	Hypersensitive response elicitor; HRE; insect resistance;
XX	
OS	biological control; transgenic plant.
XX	
XX	Erwinia amylovora.
XX	
PN	W09837752-A1.
XX	
PD	03-SEP-1998.
XX	
PF	26-FEB-1998; 98MO-US03604.
XX	
PR	28-FEB-1997; 97US-0039226.
XX	
PA	(CORR ) CORNELL RES FOUND INC.
XX	
PI	Wei Z, Zitter TA;
XX	
DR	WPI; 1998-495374/42.
XX	
DR	N-PSDB; AAV54607.
XX	
PT	use of hypersensitive response elicitor polypeptide - for
XX	
PT	application to plants or seeds or transgenic plants or seeds for the
XX	
PT	control of insects.
XX	

Erwinia chrysanthemi  
A hypersensitive r  
E. chrysanthemi hy  
Erwinia chrysanth  
Hypersensitive res  
Erwinia chrysanth  
Mycobacterium tube  
N. clavipes draglin  
Spider silk protei  
Nephila clavipes s  
N. clavipes spider  
Spider dragline va  
Mycobacterium spec  
Arabidopsis thalia  
New DNA sequence 1  
Nephila clavipes s  
Hypersensitive res  
Pseudomonas solana  
Pseudomonas solana  
Hypersensitive res  
A hypersensitive r  
Pseudomonas solana  
A hypersensitive r  
Spider dragline va  
Polymer of an anal  
Polymer of an anal  
Mouse tropinlin pr  
Arabidopsis thalia  
Spider dragline va  
Polymer of an anal  
Epstein Barr Virus  
Epstein Barr Virus  
Human tropinlin.  
Human tropinlin pr

PS Disclosure; Page 9-10; 75pp; English.

CC This is the amino acid sequence of a 39 kDa, heat stable  
 CC hypersensitive response elicitor (HRE) of *Erwinia amylovora*. The  
 CC invention relates to the use of a HRE polypeptide or protein to  
 CC control insects on plants or plants grown from seed treated with HRE.  
 CC Also claimed is a method of insect control for plants that involves:  
 CC (a) providing a transgenic plant or seed transformed with a DNA  
 CC molecule (see AAV54606-09) encoding a HRE polypeptide or protein (see  
 CC AAV75862-67); and (b) growing the transgenic plants or transgenic  
 CC plants produced from the transgenic seeds to control insects. HRE  
 CC prevents direct insect damage to plants by feeding injury. It kills  
 CC insects close to plants, and interferes with insect larval feeding  
 CC on such plants. It also prevents insects from colonising host  
 CC plants and releasing phytotoxins which result in disease damage to  
 CC plants.

CC Sequence 403 AA;

Query Match 100.0%; Score 2079; DB 19; Length 403;

Best Local Similarity 100.0%; Pred. No. 9.6e-156; Mismatches 0; Indels 0; Gaps 0;

Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSINTSGLAGSTWQISIGAGAGNNGILGTSRONAGLGGNSALGLGGGNDYVNOIAGLL 60  
 DB 1 msintsglagstwmqisigagagngilgtsrqnaglggnsalglgggnqndvtnqiaql 60  
 QY 61 TGGMMAMNMMGGGGLMGCGGLGNGLGSSGGLGEGLSNALMDMGGSLNTGSKGNN 120  
 DB 61 tggmmamnmnggggllmgcgglnglgsnglgeglalnalmggslntlgskgngn 120  
 QY 121 TTSTNSPLDOALGINSTSONDSTSGTSDSPDMOOLKMFSEIMOSLFGDGQDGT 180  
 DB 121 ttstnspldoalginstsondstsgtstdspdmoolkmfseimoslfgdgqdggt 180  
 QY 181 QGSSSGKQPTBEGBNKKGVTDAISGLMGNLSQLLNGSLGGGCGAGTGLDSSSL 240  
 DB 181 qgsssgkqptbegenkkgvtdaisglmgnglsqllngslgggcgagtgldsssl 240  
 QY 241 GGGKGLNLSGPVYQOLGNVGTGIGKAGIOALNDGTHRSSTRSFVKKGRRAAKET 300  
 DB 241 gggkglnlsgpvdyqolgnvgtgigkagioalndgthrsstrsfvkkgramakel 300  
 QY 301 GGFMDQYPEVFGKPYQKPGQEVKTDKSMARALSKPDDGWTSPASMEQFNKAKGMIKR 360  
 DB 301 ggfmdqypevfgykpyqkpgqgevtddksmaralskppddgwtspasmeqfnkakgmikr 360  
 QY 361 PMAGDTGNGMLQARGAGSSSLGIDAMMAGDAINNMALGKLGAA 403  
 DB 361 pmagdtgngnlqargagssslgidammagdainnmalgkligaa 403

RESULT 2

AAW62455 standard; Protein; 403 AA.

AAW62455;

09-NOV-1998 (first entry)

Erwinia amylovora hypersensitive response elicitor (HRE).

Hypersensitive response elicitor; HRE; growth; transgenic plant.

Erwinia amylovora.

WO9832844-A1.

30-JUL-1998.

27-JAN-1998; 98MO-US01507.

PR 27-JAN-1997; 97US-0036048.

PA (CORR ) CORNELL RES FOUND INC.

PI Beer SV, Qiu D, Wei Z;

DR WPI; 1998-427940/36.

DR N-PSDB; AAV39973.

PT Method for enhancing plant growth - comprises use of hypersensitive

PT response elicitor polypeptide or protein which may also effect, e.g.

PS increase in plant height or earlier germination seed

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Sequence 403 AA;

Query Match 100.0%; Score 2079; DB 19; Length 403;

Best Local Similarity 100.0%; Pred. No. 9.6e-156; Mismatches 0; Indels 0; Gaps 0;

Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSINTSGLAGSTWQISIGAGAGNNGILGTSRONAGLGGNSALGLGGGNDYVNOIAGLL 60  
 DB 1 msintsglagstwmqisigagagngilgtsrqnaglggnsalglgggnqndvtnqiaql 60  
 QY 61 TGGMMAMNMMGGGGLMGCGGLGNGLGSSGGLGEGLSNALMDMGGSLNTGSKGNN 120  
 DB 61 tggmmamnmnggggllmgcgglnglgsnglgeglalnalmggslntlgskgngn 120  
 QY 121 TTSTNSPLDOALGINSTSONDSTSGTSDSPDMOOLKMFSEIMOSLFGDGQDGT 180  
 DB 121 ttstnspldoalginstsondstsgtstdspdmoolkmfseimoslfgdgqdggt 180  
 QY 181 QGSSSGKQPTBEGBNKKGVTDAISGLMGNLSQLLNGSLGGGCGAGTGLDSSSL 240  
 DB 181 qgsssgkqptbegenkkgvtdaisglmgnglsqllngslgggcgagtgldsssl 240  
 QY 241 GGGKGLNLSGPVYQOLGNVGTGIGKAGIOALNDGTHRSSTRSFVKKGRRAAKET 300  
 DB 241 gggkglnlsgpvdyqolgnvgtgigkagioalndgthrsstrsfvkkgramakel 300  
 QY 301 GGFMDQYPEVFGKPYQKPGQEVKTDKSMARALSKPDDGWTSPASMEQFNKAKGMIKR 360  
 DB 301 ggfmdqypevfgykpyqkpgqgevtddksmaralskppddgwtspasmeqfnkakgmikr 360  
 QY 361 PMAGDTGNGMLQARGAGSSSLGIDAMMAGDAINNMALGKLGAA 403  
 DB 361 pmagdtgngnlqargagssslgidammagdainnmalgkligaa 403

RESULT 3

AAW61114 standard; Protein; 403 AA.

AAW61114

AAW61114;  
 26-OCT-1998 (first entry)  
 Hypersensitive response elicitor protein (39 kDa).  
 Hypersensitive response elicitor; transgenic plant; seed;  
 pathogen resistance; disease resistance; crop protection.  
 Erwinia amylovora.  
 W09824297-A1.  
 11-JUN-1998.  
 04-DEC-1997; 97WO-US22629.  
 05-DEC-1996; 96US-0033230.  
 (CORR ) CORNELL RES FOUND INC.  
 Beer SV, Qiu D, Wei Z;  
 WPI: 1998-332931/29.  
 N-PSDB; AAV36428.  
 Imparting pathogen resistance to plants. - by applying a  
 hypersensitive response elicitor polypeptide to seeds  
 Disclosure: Page 18-20; 85pp; English.  
 This is the 39 kDa hypersensitive elicitor (HRE) protein of  
 Erwinia amylovora. It is heat stable at 100 degC for at least 10  
 min, has a pI of approximately 4.3, and contains substantially no  
 cysteine. The invention relates to methods of imparting  
 hypersensitive response induced resistance to plants by treatment  
 of seeds. Isolated HRE proteins can be applied to seeds as a means  
 of imparting pathogen resistance to plants grown from the seeds.  
 Alternatively, bacteria containing the gene encoding the HRE can be  
 applied to the plant seeds, or transgenic plant seeds containing a  
 DNA molecule encoding an HRE polypeptide or protein are used. HRE  
 polypeptide sequences from Erwinia chrysanthemi, Erwinia amylovora,  
 Pseudomonas syringae, Pseudomonas solanacearum, Xanthomonas  
 campestris pv. glycines and Xanthomonas campestris pelargonii (see  
 AAW6113-18) are provided. The methods can impart pathogen  
 resistance without using agents which are harmful to the  
 environment or pathogenic to the plant seed being treated, or to  
 adjacent plants. E. amylovora causes disease in apple or pear  
 but not tomato. However, it elicits a hypersensitive response in  
 tomato. Thus, E. amylovora can be applied to tomato seeds to  
 impart pathogen resistance without causing diseases in plants of  
 that species.  
 Sequence 403 AA;  
 Query Match 100.0%; Score 2079; DB 19; Length 403;  
 Best Local Similarity 100.0%; Pred. No. 9.6e-156;  
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 MSLNTSGLAGASTWQISIGAGGNNGLGTSTRONAGIGNSALGLGGNONDTVNOIAGLIL 60  
 |||||  
 1 mslntsglagastwqisigaggnngllgtstrnaglgynsalglggnqndvtnqiaql 60  
 |||||  
 61 TGMNMMMSMGGGGGLMGGLGGGLGNGLAGSGGLGBGSNALNDMLGGLNTLGSKGCNN 120  
 |||||  
 61 tgmnmmsmggggglmgglggglgnglagsgglgbgsnaldmlgglntlgskgcnn 120  
 |||||  
 121 TTSTNSPLDQALGINSTSONDSTSGTSTSDSSDPMOQLKMFSEIMQSLFGGSDQDT 180  
 |||||  
 121 ttstnspldqalginstsondstsgtstsdssdpmoqlkmfseimqslfsgdqdt 180  
 |||||  
 181 OGSSSGGKQPTBGEDONAYKKVTDALSGLMGNLSOLLGNGLGGGGGAGTGLDGSSTL 240  
 |||||

181 gssssgkqptebgeqaykkyvtdalsgimnglsqllgngg1g9g9gnaagtgldgsstl 240  
 241 GKGKIONLSGPVDYQOGLNNAVGTGIGMKAGTIOALNDIGTHSSSTRSPVNGDRAMAKEI 300  
 |||||  
 241 gkgkionlsgpvdyqoqlnnavgtgigmkagtiolndigthssstrspvngdramakei 300  
 301 GOFMDQYPEVEFGKPOYOKPGQEVKTDDRSWAKALSKPDDGMPASMEQFNKAKGMTR 360  
 |||||  
 301 gofmdqypevfgkppyqpgqevktddkswakalskpdggmpasmeqfnkakgmtr 360  
 361 PMAGDTGNGNTQARAGGSSSLGIDAMAGDAINNALSKLGA 403  
 |||||  
 361 pmagdtgngnlgargagsslgidammagdalinna1gklga 403  
 RESULT 4  
 ID AAW87639 standard; Protein; 403 AA.  
 AAW87639;  
 09-MAR-1999 (first entry)  
 A hypersensitive response elicitor protein.  
 Hypersensitive response elicitor protein; hairpin protein;  
 disease resistance; seed quality; insect control; corn borer;  
 Lepidoptera larvae; transgenic plant.  
 Erwinia amylovora.  
 W09854214-A2.  
 03-DEC-1998.  
 28-MAY-1998; 98WO-US10874.  
 30-MAY-1997; 97US-0048109.  
 (CORR ) CORNELL RES FOUND INC.  
 (EDEN-) EDEN BIOSCIENCE CORP.  
 Beer SV, Laby RJ, Wei Z;  
 WPI: 1999-070210/06.  
 N-PSDB; AAV83989.  
 New fragments of an Erwinia hypersensitive response elicitor protein  
 and related DNA - used to impart disease resistance to plants, to  
 increase their growth and to control insects  
 Claim 4: Page 10-11; 94pp; English.  
 The present sequence represents a hypersensitive response elicitor  
 protein (also called hairpin protein) that is able to elicit a  
 hypersensitive response in plants. The specification also describes  
 hypersensitive response elicitors from other pathogenic organisms.  
 The protein, in non-infectious form, is applied to plants to impart  
 disease resistance (to a wide range of viral, bacterial and fungal  
 pathogens), to improve growth (yield, quantity and quality of seeds,  
 to provide earlier germination etc.) and to control insects (e.g. corn  
 borers, lepidoptera larvae etc.) The same results are provided by  
 transgenic plants expressing the protein.  
 Sequence 403 AA;  
 Query Match 100.0%; Score 2079; DB 20; Length 403;  
 Best Local Similarity 100.0%; Pred. No. 9.6e-156;  
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 MSLNTSGLAGASTWQISIGAGGNNGLGTSTRONAGIGNSALGLGGNONDTVNOIAGLIL 60

Query Match 100.0%; Score 2079; DB 21; Length 403;

DQ 1 msIntsglaetmgIsIggagngnllgtsrqnaIggsnsalgiygggnqndtvnqlglll 60  
| | | | |  
QY 61 TGNMAMASNMGGGGLMGGLGGLGNGLGSSGGLGEISLNLNDMLGSLNTLSKSGNN 120  
| | | | |  
DQ 61 tgmumnmnmngggImngglgglgnglgssgglgeglsalnldmlggsIntlgskgnn 120  
| | | | |  
QY 121 TTSTNSPLDQALGINSTSONDDSTSGDSNDSDDPMQOLLKMFSEIMOSTLPDGCGCT 180  
| | | | |  
DQ 121 ttstnspldqaIglnsstsnddstsgdsctsdssdpmqlllmktseImgsIlgdgqdtc 180  
| | | | |  
QY 181 QGSSSGGKOPPEEGEDONAYKKGVPTALSGLAMNGLSQLLGNGIGLGAGGNGGTGLDSSL 240  
| | | | |  
DQ 181 qgsssggkqppegegqnaykkgytaldslgImnglsqllIngglg999gngatgl dgsSI 240  
| | | | |  
QY 241 GKGGLQNTSGFVDYQQULGNNAVGTGIMKRAGIQALNDICTHRHSSTRSEFVNRGDRAAKEI 300  
| | | | |  
DQ 241 gkgglqntsgpvdYqqlgnavgtlgmkagIqalIndigtHrHsstrsfvnrkdramakeI 300  
| | | | |  
QY 301 GGFMDQYDEVEVGKXOYOXGPRQEVKTDTDKSMAKALKRPDDDGMPHPASBQNKAKGMTR 360  
| | | | |  
DQ 301 ggfmdqyevvgkxpyqypqvewtkdkswakalskpdddgmpasmeqInkakgmIkr 360  
| | | | |  
QY 361 PMAGDTGCNLTQARGAGSSSLGIIDAPMAPADAIINNALGKLGA 403  
| | | | |  
DQ 361 pmagdtgcngltargaggssIgidammagdaiinmalgklga 403  
| | | | |

RESULT 5  
AAV71093  
ID AAV71093 standard; Protein: 403 AA.  
XX  
AC AAV71093:  
XX  
DT 08-SEP-2000 (first entry)  
XX  
DE Erwinia amylovora hypersensitive response elicitor #1.  
XX  
KW Hypersensitive response elicitor; environmental stress resistance;  
KM plant.  
XX  
OS Erwinia amylovora.  
XX  
PN W0200028055-A2.  
XX  
PD 18-MAY-2000.  
XX  
PF 04-NOV-1999; 99W0-US26039.  
XX  
PR 05-NOV-1998; 98US-0107243.  
XX  
PA (EDEN-) EDEN BIOSCIENCE CORP.  
XX  
PI Wei Z, Schading RL;  
DR WPI: 2000-3/6566/32.  
DR N-PADB; AAD00668.  
XX  
PT Application of a hypersensitive response elicitor protein to plants to  
impart stress resistance -  
PS Disclosure; Page 7-8; 84pp; English.  
XX  
CC The patent discloses a method to impart stress resistance to plants by  
applying a hypersensitive response elicitor in a non-infectious form to  
a plant or seed. The present sequence is a hypersensitive  
CC response elicitor protein from Erwinia amylovora. It is  
CC used to impart stress resistance to plants.

	Best Local similarity    100.0%; Fred. No. 9.6e-156;					
	Matches	403;	Conservative	0;	Mismatches	0; Indels
						Gaps
						0;
QY	1	MSLNTSGIGASTMOISIGAGGNNGLTSTSRONMGLGSNSALGELGGNQNDTVQLACL	60			
Dd	1	mSLntsgIgaStmqIsIgaggnngllgtsrqnaqlggnsalglggndqdtvqIagll	60			
QY	61	TGNMMAMSMWGCGGLMGGLGGGNGCNGSGLGSSGCLGEICSNALNMDLGSLWLTLSKGGNN	120			
Dd	61	tGnmAmmsmWGggImgggslggylgnqlgssggylgeglsalnmdlpgsIntlgskggnn	120			
QY	121	TTSTPNSPLDQALGINSTRSONDDTSRGTDRSDSPMQOOLLKMPSLIMOSLFQGGCPGT	180			
Dd	121	tTstpnSpldqAlginstrsonddtsrgtDrSdsdpmqoollkmPsLI MOSLf qdgcpgt	180			
QY	181	OQSSESGROPTEEGEONAYKKGVTDALSGILMNGLSQLGNGSLGGQGGMATGIDXSLL	240			
Dd	181	qgsseSgRpTeegEnaykkytDALsgIlmgnglsqllgnggylggggngagtIdgxSl	240			
QY	241	GKGKLQNLSGPVDYQOLGNNAVGTIGMKACIOALNDICTHRHSITRSVYNKDRAAKEI	300			
Dd	241	gKgklqnlsGpvdyqqlgnavgtIgmkaqiAldictHrsItRsFynkdReakeI	300			
QY	301	GOFMQYEVEVGKPXYOGPGQEVETDDKSNAKALSKRPDDGMPASMEQFWKAKMTKR	360			
Dd	301	gfMqyeVeVGkpYogPgqeVetddkswKaIskpoddmpasmeqfInkakmtkr	360			
QY	361	PMAGPTGNGNLQARAGASSLGDIMAMAGDAINNALIKRLGA	403			
Dd	361	pmaGptgnGlqarAgasslgidamagdaInnalIkrlGa	403			
RESULT	6					
ID	AAY84854					
	AAY84854 standard; Protein: 403 AA.					
XX						
AC	AAY84854;					
XX						
DT	08-AUG-2000 (first entry)					
XX						
DE	A hypersensitive response elicitor protein.					
XX						
KW	Hypersensitive response; insect control; disease resistance;					
KM	hypersensitive response elicitor; plant growth; vegetable; crop;					
KW	ornamental plant.					
XX						
OS	Ervinia amylovora.					
XX						
Key	Location/Qualifiers					
FH	Peptide	169..403				
FT	/note= "claimed under claim 5"					
FT	Peptide	210..403				
FT	/note= "claimed under claim 5"					
FT	Peptide	267..403				
FT	/note= "claimed under claim 5"					
FT	Peptide	343..403				
FT	/note= "claimed under claim 5"					
FT	Peptide	105..179				
FT	/note= "claimed under claim 7"					
FT	Peptide	121..150				
FT	/note= "claimed under claim 7"					
FT	Peptide	137..166				
FT	/note= "claimed under claim 7"					
FT	Peptide	137..156				
FT	/note= "claimed under claim 7"xX					
XX						
PN	WO200020452-A2.					
XX						
PD	13-APR-2000.					
XX						
PF	05-OCT-1999;	99WO-US23181.				
XX						
PR	05-OCT-1998;	98US-0103050.				

XX (EDEN-) EDEN BIOSCIENCE CORP.  
 XX  
 XX  
 PI Mel Z, Fan H, Nigemeyer Jr;  
 XX  
 DR WPI, 2000-303745/26.  
 XX N-PSDB; AAL4938.  
 PT Hypersensitive response elicitor polypeptides useful for imparting  
 PT enhanced growth, disease resistance and insect resistance to plants,  
 PT especially vegetables and ornamental flowers -  
 PS  
 PS Claim 4; Page 8-10; 100pp; English.  
 CC The present sequence represents a hypersensitive response elicitor  
 CC polypeptide. The specification describes hypersensitive response  
 CC elicitor polypeptide fragments, which do not elicit a hypersensitive  
 CC response. Instead, the proteins impart disease resistance to plants,  
 CC enhance plant growth, and/or control insects. The polypeptide  
 CC fragments may be used to these properties to plants. The plants which  
 CC may be treated in this way include vegetables, crops and ornamental  
 CC plants such as alfalfa, rice, wheat, barley, rye, cotton, sunflower,  
 CC peanut, corn, potato, sweet potato, bean, pea, chickory, lettuce,  
 CC endive, cabbage, brussel sprout, beet, parsnip, turnip, cauliflower,  
 CC broccoli, radish, spinach, onion, garlic, eggplant, pepper, celery,  
 CC carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon,  
 CC citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,  
 CC tomato, sorghum or sugarcane, Arabidopsis thaliana, Saintpaulia,  
 CC petunia, pelargonium, poinsettia, chrysanthemum, carnation or zinnia.  
 XX  
 XX Sequence 403 AA;  
 SQ  
 Query Match 100.0%; Score 2079; DB 21; Length 403;  
 Best Local Similarity 100.0%; Pred. No. 9.6e-156;  
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSLSGIGASTMIOISIGAGGNNGLTSTRONAGIGNSALGIGGQNDVTNOLAGIL 60  
 Db 1 msltsigastmqlsigsagngmlqtsrqnaqlgnsalglggnqndvtnqlagll 60  
 QY 61 TGMAMMSMNGGGLMGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGL 120  
 Db 61 tgmammsmngggglmgggllggglggglggglggglggglggglggglggglgggn 120  
 QY 121 TSTSTNSPLDQALGINSNDSTSGTSDTSDSPMOQLKMSSEIMQSIFGGGOGGT 180  
 Db 121 tststnspldqlglnstsqndstsgtscdsdsdpmqllkmtselmqslfsgdgggt 180  
 QY 181 QGSSSGKQPTGEGEONAKKGVTDALSGLMGNGLSQLGNGGLGAGGOGGAGTGLDSSL 240  
 Db 181 qgsssgkqptegegonakkytvdalsglmgnglsqllgngglgagggngagtgldgssl 240  
 QY 241 GKGGLQNTSGPVDYQOLGNAVGITGMRAGIOALNDIGTHRSSTRSFVNKGDRAAKEI 300  
 Db 241 gkgglqntsgpvdqyqlgnavgitgmkagiqalndigthrsstrsfvnkgdrakei 300  
 QY 301 GGFMOQYEVFEGKPOYOKPGQEVKTDKSWAKALSKPDDOCTPASMOPFKAKGMIR 360  
 Db 301 ggfmoqyevfegkpoqokpgqevktddkswakalskppddqmtcpasmeqfinkakgmir 360  
 QY 361 PMAGDTGNGLQARGAGSSSLGIDAMAGDAINNALGKLGAA 403  
 Db 361 pmaqdtgnnglqargagssslgidamagdaalnnaigkligaa 403  
 RESULT 7  
 AAM06598  
 ID AAM06598 standard; Protein; 385 AA.  
 AC AAM06598;  
 XX  
 DT 30-MAR-1997 (first entry)

XX  
 DE Hypersensitive response elicitor protein.  
 XX  
 KW Hypersensitive response; elicitor; Erwinia amylovora; plant;  
 KW disease-resistance; Escherichia coli; infiltration; virus;  
 KW bacterium; fungus; pathogen; biological control agent.  
 XX  
 OS Erwinia amylovora.  
 PN WC0639802-A1.  
 XX  
 PD 19-DEC-1996.  
 XX  
 PF 05-JUN-1996; 96MO-US08819.  
 XX  
 PR 07-JUN-1995; 95US-0475775.  
 XX  
 PA (CORR ) CORNELL RES FOUND INC.  
 XX  
 PI Beer SV, Wei Z;  
 XX  
 DR WPI: 1997-051614/05.  
 DR N-PSDB; AAT49314.  
 PT Imparting pathogen resistance to plants - with hypersensitive  
 PT response elicitor polypeptide or protein  
 XX  
 PS Claim 7; Page 46-47; 69pp; English.  
 CC This sequence represents a hypersensitive response elicitor from  
 CC Erwinia amylovora, with a mol.wt. of 37,000. The elicitor has a pI  
 CC of 4.3, thermostability at 100 deg C for at least 10 min, and  
 CC contains no cysteine. The elicitor may be used in a new method for  
 CC imparting pathogen resistance to plants, by application of the  
 CC elicitor in a non-infectious form to plant cells, by spraying,  
 CC infection, leaf abrasion, or plant infection with recombinant  
 CC bacteria (non-infectious to the host plant, e.g. Escherichia coli)  
 CC expressing the elicitor as a biological control agent, to allow  
 CC recombinant protein infiltration into the plant. The method  
 CC confers virus, bacterium or fungus disease-resistance on crops and  
 CC ornamental plants.  
 XX  
 XX Sequence 385 AA;  
 SQ  
 Query Match 92.7%; Score 1928; DB 18; Length 385;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-144;  
 Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSLSGIGASTMIOISIGAGGNNGLTSTRONAGIGNSALGIGGQNDVTNOLAGIL 60  
 Db 1 msltsigastmqlsigsagngmlqtsrqnaqlgnsalglggnqndvtnqlagll 60  
 QY 61 TGMAMMSMNGGGLMGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGL 120  
 Db 61 tgmammsmngggglmgggllggglggglggglggglggglggglggglggglgggn 120  
 QY 121 TSTSTNSPLDQALGINSNDSTSGTSDTSDSPMOQLKMSSEIMQSIFGGGOGGT 180  
 Db 121 tststnspldqlglnstsqndstsgtscdsdsdpmqllkmtselmqslfsgdgggt 180  
 QY 181 QGSSSGKQPTGEGEONAKKGVTDALSGLMGNGLSQLGNGGLGAGGOGGAGTGLDSSL 240  
 Db 181 qgsssgkqptegegonakkytvdalsglmgnglsqllgngglgagggngagtgldgssl 240  
 QY 241 GKGGLQNTSGPVDYQOLGNAVGITGMRAGIOALNDIGTHRSSTRSFVNKGDRAAKEI 300  
 Db 241 gkgglqntsgpvdqyqlgnavgitgmkagiqalndigthrsstrsfvnkgdrakei 300  
 QY 301 GGFMOQYEVFEGKPOYOKPGQEVKTDKSWAKALSKPDDOCTPASMOPFKAKGMIR 360  
 Db 301 ggfmoqyevfegkpoqokpgqevktddkswakalskppddqmtcpasmeqfinkakgmir 360



OY	13	MOIST- 1 mqlltckahlgddlqvsqldga-	GGAGGNNDLIGTSRNNALGSG-NSA- gkglinsaaslsqsvskltslclsaletsmm---	---ISGLGGQNQNTFYVQNLGLITGMAMM	67
Dd					55
OY	68	SMWGCGGIMGGGLGGGJGNGJGSGGJGLEGLSNALNMLGSLTLTKSKGGNNTTSTPINS			127
Dd	56	-----fggalagqlgas-skyqlmsnqlqgsfqn-----			96
OY	128	PLDQALGINSTSDNDSTGSTDTSTDSDDPMQQLLKMFSEILOSFG- 97 -----sgydals-----kmfdkalddlghdtvrlktngsnq			128
OY	180	TQGSSSGKOPTBEGDONAYKKGYVDALSGILMNGLLSQLLGNGJGSGGOGGAAGCLDGSS			235
Dd	129	lanmlnaasqnlctgmmafigesgvnalssllgnljgasm-----sfqsps			174
OY	240	LGGKGLOMLSGRPVVYQQLGNAVGVGIGKKAGIQALINDIGTHRHSSSTRSFYNKKCDRAAKE			299
Dd	175	Ilgagdlgglsqagafnqlgnalimgvgvgnalaalslnvsthdvgnmhfydkedrgmake			234
OY	300	IGGFMDQYPEVFGRPOYQKGGEGEYKTDKSWAKALSKEPDDEGTPLASMEQFNKAQMNIK			355
Dd	225	Iqglmndypelrfgyrpxykdqswsspktdckawakalskdpddntgtasmdkfrrqamgmik			294
OY	360	RPMAGDTGNGMLQARGAGSSLSIGIDAMMAGDATNNMALGRIGAA			403
Dd	295	savgdctgntlnlrigaggasilidaavaavgdkianmslqklana			338
RESULT	10				
ID	AAM87638				
AC	AAM87638 standard; Protein; 338 AA.				
XX					
AC	AAM87638;				
XX					
DT	09-MAR-1999 (first entry)				
XX					
DE	A hypersensitive response elicitor protein.				
KW	Hypersensitive response elicitor protein; hairylin protein; KM disease resistance; seed quality; insect control; corn borer; KW Lepidoptera larvae; transgenic plant.				
XX					
OS	Erwinia chrysanthemi.				
XX					
PN	M09854214-A2.				
PD	03-DEC-1998.				
PF	28-MAY-1998; 98WO-US10874.				
XX	30-MAY-1997; 97US-0048109.				
RR	(CORR.) CORNELL RES FOUND INC. PA (EDEN-) EDEN BIOSCIENCE CORP.				
PI	Beer SV, Laby RJ, Wei Z;				
DR	WPI; 1999-070210/06.				
DR	N-PSDB; AAV835988.				
PT	New fragments of an Erwinia hypersensitive response elicitor protein and related DNA - used to impart disease resistance to plants, to increase their growth and to control insects				
XX					
P5	Disclosure; Page 7-8; 94pp; English.				
CC	The present sequence represents a hypersensitive response elicitor protein (also called hairylin protein) that is able to elicit a hypersensitive response in plants. The specification also describes hypersensitive response elicitors from other pathogenic organisms. The protein, in non-infectious form, is applied to plants to impart				

CC disease resistance (to a wide range of viral, bacterial and fungal  
CC pathogens), to improve growth (yield, quantity and quality of seeds,  
CC to provide earlier germination etc.) and to control insects (e.g. corn  
CC borers, Lepidoptera larvae etc.) The same results are provided by  
CC transgenic plants expressing the protein.

xx Sequence 338 AA:

SO

Query Match 34.6%; Score 718.5; DB 20; Length 338;  
Best Local Similarity 42.8%; Pred. No. 6.9e-49;  
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;

OY 13 MOIST-GGAGCNGNGLIGTSRONAGLGG-NSA---LGISGONONTVNOLAGLLTGMAMMM 67  
DB 1 mgtltkahlgdlgyvsjlyaq--gltklnsaaaslgssvdkltsctldkltalsltmm--- 55  
OY 68 SMMGGGGLMGGGLGGLGNGLCGGSGGLEEGSLBNALNDMLGSLNLTLSKGNNTSTNS 127  
DB 56 -----fgyalagqglyas-skglgmnglqgsfgn----gaqgasnllyspk- 96  
OY 128 PLDQALGINSTQNDSGTSDTSDDSPMOQLKMFSEIMQSIFG-----DGODG 179  
DB 97 -----sgdals-----kmfdkalddllghndtvcklknqsg : 128  
OY 180 TQGSSSGGKOPTGEHOMAYKKGVTDALSGLMNGLSQLGNCGLGGCGGAGTAGLDSSS 239  
DB 129 lamslnasgmtqumnaifsgynmlsalnlgnylqgsm-----sgfsqps 174  
OY 240 LGGKGLONLSGVDVOQOLCNAYGTGICMKAGIQIALINDICTHRHSSTRSPVNGDRAMKE 299  
DB 175 lgaagllqgisgagatnqlgnalnigmvgvquaalasalnsvtlhvdgnrhtvdekdrfmake 234  
OY 300 IOGFMDQFEVEFGKPOYOXKPGCEWKTDDKSWAKALSFKPPDDGMTPASMEOPNKAKGMIR 359  
DB 233 lsgfmdqypelfghkpeyqdgdwsspktdkksvakalslpdddgmtygaandkfrtgamgnlk 294  
OY 360 RPMAGDTGNGNLQARAGSGSSLGIDPAMAADAINNNALGKLGA 403  
DB 295 savagdtgtntlnlrqagagsligdaavvydklanmslqklana 338

RESULT 11  
AAW82407  
ID AAW82407 standard; Protein: 338 AA.  
XX  
AC AAW82407;  
XX  
DT 23-FEB-1999 (first entry)  
XX  
DE  
KM E. chrysanthemi HrpN-ECH protein.  
XX  
KW Hypersensitive response elicitor protein; hrpn-Ech; pathogen resistance;  
XX plact; transformation; pathogen-inducible promoter.  
OS Erwinia chrysanthemi.  
XX  
PN US5650015-A.  
XX  
PD 15-DEC-1998.  
XX  
PF 07-JUN-1995; 95US-0484358.  
XX  
PR 07-JUN-1995; 95US-0484358.  
XX  
PA (CORR ) CORNELL RES FOUND INC.  
XX  
PI Bauer D, Collmer A;  
XX  
DR WPI: 1999-069852/06.  
DR N-PSDB: AAV73494, AAV73507.  
XX  
PT DNA encoding Erwinia chrysanthemi hypersensitive response protein



PT hrpN - useful for imparting pathogen resistance to plants  
 XX  
 PS Claim 2, Column 29-30; 27pp; English.

CC This sequence represents a novel Erwinia chrysanthemi protein, hrpN-Ech,  
 CC that elicits a hypersensitive response in plants. The encoding DNA can be  
 CC used for imparting pathogen resistance to plants, by transforming a  
 CC plant with a vector containing the DNA and a pathogen-inducible promoter.  
 XX

SO Sequence 338 AA:

Query Match 34.6%; Score 718.5; DB 20; Length 338;

Best Local Similarity 42.8%; Pred. No. 6.9e-49;

Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;

QY 13 MQIST-GGAGGNNGLGTSRONAGLGG-NSA---LGIGGNDQVYNQLAGLITGMMMM 67  
 DB 1 mqitlkahlggdlgvsqglaq--gikylnsaaslgssvdklsldkltalsltmm--- 55

QY 68 SMWGGGLMGGLGGGIGNGIGSGGLGEGLSNALNDMLGSLNTLGSKGNNTTSTNS 127

DB 56 -----fgalaaglgas-skylgmsnqlgsgfign-----gaqasnllyvpk- 96

QY 128 PLDQALGINSTSONDSTGCTDSTSDSPMQQLKMFESIMSLFG-----DGQDG 179

DB 97 -----sggdals-----kmfdaldldllghdvtkltnqsnq 128

QY 180 TQSSSGGKQPTGEONAYKKGYTDALSGLMGNGLSQLGNGIGGGGNGAGTGLDSS 239

DB 129 lamslnasqmtgymnaafgsvynalssllnglqgsm-----sgfsqps 174

QY 240 LGSGGLQNTSLGPDYDQOLGNAVGTGIGMKAGIOALNDIGTHRSSTPSFVNKGDRAVAK 299

DB 175 lggsgylgslgagafnglgnalimgvqnaaalsalnsvthvgdnhrhvkdeirgmake 234

QY 300 IGQFMDOYPEVFGKPOYOKPGQGEVKTDDKSMKALSKPDDGNTPASMQFKNKAKMIK 359

DB 235 lggfmdqypelfgkpeyqkqgwsppktdkswakalskpdddgmtgasmckfrqangmik 294

QY 360 RPNAGDTGNGNLQARGAGSSLGIDAMAGDAINNALGKLGAA 403

DB 295 savagdtgntlnlrgagsgslgidaavvgdklamslglkana 338

RESULT 12

AA71092  
 ID AAY71092 standard; Protein; 338 AA.

XX AAY71092;

08-SEP-2000 (first entry)

DE Erwinia chrysanthemi hypersensitive response elicitor.

XX Hypersensitive response elicitor; environmental stress resistance;

KW plant.

XX Erwinia chrysanthemi.

OS Key Location/Qualifiers

FT MISC-difference 19 /note- "Encoded by GCGCTGGGT"

XX MO200028055-A2.

PD 18-MAY-2000.

PF 04-NOV-1999; 99MO-US26039.

XX 05-NOV-1998; 98US-0107243.

XX (EDEN-) EDEN BIOSCIENCE CORP.

XX  
 PI Wei Z, Schading RL;

XX WPI; 2000-376566/32.

DR N-PSDB; MADO00667.

PT Application of a hypersensitive response elicitor protein to plants to

PS Import stress resistance

XX Disclosure; Page 4-5; 84pp; English.

CC The patent discloses a method to impart stress resistance to plants by  
 CC applying a hypersensitive response elicitor in a non-infectious form to  
 CC a plant or seed. The present sequence is a hypersensitive  
 CC response elicitor protein from Erwinia chrysanthemi. The protein is heat  
 CC stable and used to impart stress resistance to plants.  
 XX

SO Sequence 338 AA:

Query Match 34.6%; Score 718.5; DB 21; Length 338;

Best Local Similarity 42.8%; Pred. No. 6.9e-49;

Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;

QY 13 MQIST-GGAGGNNGLGTSRONAGLGG-NSA---LGIGGNDQVYNQLAGLITGMMMM 67

DB 1 mqitlkahlggdlgvsqglaq--gikylnsaaslgssvdklsldkltalsltmm--- 55

QY 68 SMWGGGLMGGLGGGIGNGIGSGGLGEGLSNALNDMLGSLNTLGSKGNNTTSTNS 127

DB 56 -----fgalaaglgas-skylgmsnqlgsgfign-----gaqasnllyvpk- 96

QY 128 PLDQALGINSTSONDSTGCTDSTSDSPMQQLKMFESIMSLFG-----DGQDG 179

DB 97 -----sggdals-----kmfdaldldllghdvtkltnqsnq 128

QY 180 TQSSSGGKQPTGEONAYKKGYTDALSGLMGNGLSQLGNGIGGGGNGAGTGLDSS 239

DB 129 lamslnasqmtgymnaafgsvynalssllnglqgsm-----sgfsqps 174

QY 240 LGSGGLQNTSLGPDYDQOLGNAVGTGIGMKAGIOALNDIGTHRSSTPSFVNKGDRAVAK 299

DB 175 lggsgylgslgagafnglgnalimgvqnaaalsalnsvthvgdnhrhvkdeirgmake 234

QY 300 IGQFMDOYPEVFGKPOYOKPGQGEVKTDDKSMKALSKPDDGNTPASMQFKNKAKMIK 359

DB 235 lggfmdqypelfgkpeyqkqgwsppktdkswakalskpdddgmtgasmckfrqangmik 294

QY 360 RPNAGDTGNGNLQARGAGSSLGIDAMAGDAINNALGKLGAA 403

DB 295 savagdtgntlnlrgagsgslgidaavvgdklamslglkana 338

RESULT 13

AA784853  
 ID AAY84853 standard; Protein; 338 AA.

XX AAY84853;

08-AUG-2000 (first entry)

DE A hypersensitive response elicitor protein.

XX Hypersensitive response; insect control; disease resistance;

KW hypersensitive response elicitor; plant growth; vegetable; crop;

XX ornamental plant.

OS Erwinia chrysanthemi.

XX Key Location/Qualifiers

FT MISC-difference 20 /note- "encoded by CTGGCTGCT"



Db	235	1gqfmdqypelfgpeyqkdgkwssprkcdkdxswakalshkppddqmtgsmndkfrqangmik	294
QY	360	RPMAGDGTGNGMLQARAGAGSSLGIDAMMAGDAINNMALGKLGAA	403
Db	295	savagdtcgtntlnlrqgagsalqdaavvgdklanmlqykiana	338
RESULT	15		
AAW75862			
ID	AAW75862	standard; Protein; 340 AA.	
AC	AAW75862;		
XX			
DT	07-DEC-1998	(first entry)	
DE	Erwinia chrysanthemi hypersensitive response elicitor (HRE).		
XX			
KM	Hypersensitive response elicitor; HRE; insect resistance;		
KM	biological control; transgenic plant.		
XX			
OS	Erwinia chrysanthemi.		
XA			
Key	Location/Qualifiers		
FT	Misc-difference 20		
FT	/note="Leu-20 residue deduced from nucleotide		
FT	sequence in not included in the HRE		
FT	provided in the specification"		
FT	Misc-difference 21		
FT	/note="Gly-21 residue deduced from nucleotide		
FT	sequence in not included in the HRE		
XX	provided in the specification"		
XX			
PN	NO9837752-A1.		
PD			
PD	03-SEP-1998.		
XX			
PF	26-FEB-1998; 98MO-US03604.		
XX			
PR	28-FEB-1997; 97US-0039226.		
XX			
PA	(CORR ) CORNELL RES FOUND INC.		
PI			
PI	Wei Z, Zitter TA;		
DR	WPI: 1998-405374/42.		
DR	N-PSDB; AAW54606.		
XX			
PT	Use of hypersensitive response elicitor polypeptide - for		
PT	application to plants or seeds or transgenic plants or seeds for the		
PT	control of insects.		
PS	Disclosure; Page 7-8; 75pp; English.		
XX			
CC	This is the amino acid sequence of a 34 kDa, heat stable		
CC	hypersensitive response elicitor (HRE) of <i>Erwinia chrysanthemi</i> .		
CC	The invention relates to the use of a HRE polypeptide or protein to		
CC	control insects on plants or plants grown from seed treated with HRE.		
CC	(a) Claimed is a method of insect control for plants that involves:		
CC	providing a transgenic plant or seed transformed with a DNA		
CC	molecule (see AAW54606-09) encoding a HRE polypeptide or protein (see		
CC	AAW75862-67); and (b) growing the transgenic plants or transgenic		
CC	plants produced from the transgenic seeds to control insects.		
CC	HRE prevents direct insect damage to plants by feeding injury. It kills		
CC	insects close to plants, and interferes with insect larval feeding		
CC	on such plants. It also prevents insects from colonising host		
CC	plants and releasing phytotoxins which result in disease damage to		
XX	plants.		
Sequence	340 AA;		

```

Best Local Similarity 42.6%, Pred. No. 6.9e-49,
Matches 172; Conservative 41; Mismatches 114; Indels 77; Gaps 10.

QY 13 MOIS1-GGAGGNNGLIGTSSRONAGIG-NSA---LGLGGGNDVTYNOLAGLITGMMMA 67
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1 mgtltkahlggdllysgyglgagqjklgjnlnsaasslygsvdrklsrtldkltsaaltmm---- 57

QY 68 SMGGGGLMGGLGGGGLGNGLCGSGGLGEGLSNALNDMLGSSLNTLGSKGKNNTSTTNS 127
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 58 -----fggalagglgas-skjlysmnqlygsfgn-----gaqgaanllsyk- 98

QY 128 PLDDALGINSTSQNDNSTGTSSTSDSSDPMQQLKMFSEIMQSLEF-----DGQDG 179
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 99 -----sggdals-----kmtkaldldlghdtvckltlqgsq 130

QY 180 TQSSSGGKQPEEGEBOQNAKKGTVDALSGLMGNGLSQLGNGGLGGGCGGNAGTGIDGSS 239
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 131 lamslnasqmtqgmnaafsgymnaalsllyngqlygsm-----sgfsqps 176

QY 240 LGGKGLQNLGCEPVDYQOLGNAVGSTGIMKAGIQIALNDICTHRRHSSTRFVKNKDRAMAKE 299
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 177 lgaagllglsygafenqjlnaigvmgygnaalsnvaethvdmnrhrtvkedrgmake 236

QY 300 IGGFMDQPEVYNGKFOYQKGPQGEVKTDDKSNAAKALSKRPDDGCMPPASMEQFNKAKGMK 359
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 237 lggfmndqypellfgkpeyqkdgywsapktdtdkswakalskppddmgfysamndkfrgmgmik 296

QY 360 RPMAGDTGNGNLQARGAGSSSLGIDAPMAAGDILNNNALGKLGA 403
      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 297 savagdgntctnlrtgaggaslgidaaavvgdklamsysklana 340

```

Search completed: October 25, 2001, 11:43:56  
Job time: 32 sec











QY	94	LGEGLSNALNDMLGGSLINTLGSKSGNNMTSTSTNPLDPAALGINSTQNDSDSTGTDSD	153
Dd	355	AG-GILALGHAACATATPPSGNGSGDGCIGATANSP-L--QAGAGGNG--GHGLVNGGTCG	411
QY	154	SSDEWQQLLKPFSEIMOSLFGDGODGTGSGSSGKOPFEGEQNAKRYKGVTDALSGLMNG	213
Dd	412	AG-----GAGHAGSTGATGTLAQTPGGNCT---NGAGGHHGNGNG	450
QY	214	LSQLLNGGLG-----GGGCGNAGTGLDGSSLAGKGLQNLNSGPYDQOLNAAVGTGICM	268
Dd	451	GAQ-HGGGCGVGKGGAGGSSGAGGNGFPAATLSPBGADCGNG-----GNGKGGGDDGK	502
QY	269	AGIQALNDIGTHRRSSTRSPVNNKGDRAMAKEIIGQFMQYPEVFGRKPOXQKPGGEVKTDD	328
Dd	503	AG-----DGGAGAAGDVTYLAIVNGAGG-----DGGNGGEVTVGG	536
QY	329	KSMAKALS-----KPPDDGMTPASMEQFNKAKGMIKRPAAGTCN-----GN	370
Dd	537	KGGAGGVSANPALNGSACANGTAPTSGCGNGCGAGATPTVAEGENGAGCGNGHGGSVGN	596
QY	371	LQARGAGSSSL---GIDAMAGDAIINNMAIGKIGAA	403
Dd	597	GGAGGAGGNGVAGTCL--ALNGGNGGNGGIGGNGGSA	631

RESULT 10  
 A70869  
 Hypothetical glycine-rich protein RV2490c - Mycobacterium tuberculosis (strain H37Rv)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C:Accession: A70869  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajadurai, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: A70869  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1660 <OL>  
 A:Cross-references: GB:AL021246; GB:AL123456; NID:g3261507; PIDN:CAA16067.1; PID:g2791555  
 A:Experimental source: strain H37Rv  
 C:Genetics:  
 A:Gene: RV2490c  
 C:Superfamily: collagen alpha 1(IV) chain

	Query Match	11.8%	Score 245.5;	DB: 2;	Length 1660;
	Best Local Similarity	28.0%;	Pred. No. 7.7e-08;		
	Matches 115;	Conservative 21;	Mismatches 159;	Indels 115;	Gaps 20;
OY	7	GLGASTMQISIGCAGCGNNGLLGTSTRQNGILGNSALGLGGCQNQNTVWQLAGLTLGNMNM	66		
		: :			
Db	142	CNGAAGCAVGQVGGACGAAGLFGI---	GCAGCAGCAGAPGT--GGTGMWLG-----	188	
OY	67	MSMGGGGILMGSLGGLGGLN-----	GLGSGSLGEGLSNALNTMLGSLTLGSKGN	119	
		:			
Db	189	GGGVGGMGAGGAGGAGAGSNAGLPENGCGACGAGAGGAGAGA----	GGNAGMPFHGGAG	243	
OY	120	NTTSTTNPLDQALINSTSQNDSTSGTSDTSDDSPMQOLKMFSEIMOSLFPDGODG	179		
Db	244	GVGGV-----CAACANGATPCQDCAGAVASDDAG-----	GDGLG	280	
OY	180	TGGSSSGCKQPTEGEONAYKKCVTDALSGLMNG--	LSOLLGNGLG--GGGGNAGTGL	235	
		:			
Db	281	SIDGGGGA-----	GGVCGNGRGRCILLNCGAGGAGGVGAGGAGAGA	321	
OY	236	DSSSLGKGGLNLGPVDYQQQLNAVCT---	CIEMKAGICIALNDITGRHSSTSPYAK	291	
		:			
Db	322	AGCA--GGAGATGINPAGISAAGGAGGAGCAGCAGNGGAGGAGAG-----		367	

QY	292	GDRAAAKKEIGEDMDY	PEVFFK	POYQKQKPGGEVNTD	TKSNAAKALSKRPDDGDMTPASMEQF	351
Db	368	GSAGILGTYVGRAGDGGAGGGGGLGAPD	-G	-GAGNGSGTWLA	-----	413
QY	352	NKAKMTIRPMAAGDTGNGCNLQARAGSGSSLG	TDAMMA	GDAITNNAL	KLG	401
Db	414	-AGG- - - - -	HGGDPGLG	-AGGAGGSGGAGA	RAG- - - - -	ANGLAAGDGG
						451

RESULT 11  
 E70820  
 hypothetical glycine-rich protein RV0977 - *Mycobacterium tuberculosis* (strain H37RV)  
 C:Species: *Mycobacterium tuberculosis*  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C:Accession: E70820  
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno  
 A:Reference number: AF0500; MUID:98295987  
 A:Accession: E70820  
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-923 <COL>  
 A:Cross-references: GB:AL021999; GB:AL123456; NID:g3261538; PIDN:CAA17576.1; PID:g2921  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: RV0977  
 C:Superfamily: elastin

```

Query Match 11.8%; Score 245; DB 2; Length 923;
Best Local Similarity 27.1%; Pred. No. 4.2e-08;
Matches 114; Conservative 30; Mismatches 166; Indels 110; Gaps 19;

QY      5 TSGLGASTMQISIGGAGCNGNGLGTSRONAGLGSNSALGLGGGNONDVNTVLQAGILTTGM 64
          :|:|: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      198 SGIIGA-----GGAGMGGLLPFGNGAGAGGGSSGGLGSGCT----- 235

QY      65 MMMSWGGGGLMCGGIG--GIGL--NLGGSGGICEGLSNALNDMLG---GSLNTTGSK 116
          ||:|: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      236 -----GGAGMAAGPAGGGTGIGTIGTAGAGVG-GHSAFLFHGGINGDGTGMGQG 288

QY      117 GGNNITSTTNPLDQALGINSTSQNDNSTSRTSDTSDDPMQLKKFSEIMQSIPEDG 176
          ||:|: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      289 GAGAGGMAAEIITVGIEGGGGGDGAGAGGIGTSSGIGG-----SQAGAHG 340

QY      177 OGTQTSSSGGKOPTGEQNAYKKGVTDALSGLMNG--LSOLLKGSLG--GGGNGAGT 233
          ||:|: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      341 GGGGGGAGSGSGVGGGAGAGDGAGGIGTGTGNGSIGGAAGCNGCRGAGGMTA 400

QY      234 GLDSSLGKGLQNLGSPVDYQOLGNVGT--GIGMKAGIQALINDIGTHHSSTRSFVKNG 292
          ||:|: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      401 GSDGCGGGGG-----NGGVGSAGAGGTGDDGGAAG-----GAPGH----- 441

QY      293 DRAMKEIQPDQRYEVFGKPYOKPGQEYKTTDDKSMAKLSPRDDGMPASMEQFN 352
          ||:|: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      442 -----GYFOOPAO--GLPIGTGTGE-----GGAGGAGDGGOGDIG--FD 480

QY      353 KAKGMIKRP-----MAGDTNGNLDAR-----GAGGS-----SLGIDAMAGDA 391
          :|:|: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      481 GGRGGGCGGGGGGAGGDSSTFPNAQANNNGDGGAGVGAGAGTGTGTGVCADGGRGDS 540

RESULT 12
F70963
hypothetical glycine-rich protein Rv2634c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Nov-1999
C:/Accession: F70963

```



R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
 A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A: Reference number: A70500; MUID: 98295987  
 A: Accession: F70963  
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A: Molecule type: DNA  
 A: Residues: 1-778 <COIL>  
 A: Cross-references: GB:280225; GB:AL123456; NID:93242265; PIDN: CAB02341.1; PID: e266390;  
 A: Experimental source: strain H37RV  
 C: Genetics:  
 A: Gene: RV2634C  
 C: Superfamily: unassigned collagens

Query Match 11.6%; Score 242; DB 2; Length 778;  
 Best Local Similarity 30.1%; Pred. No. 5.3e-08;  
 Matches 91; Conservative 19; Mismatches 110; Indels 82; Gaps 14;

QY 7 GCGASPMQISIGAG--GNNGLGTSRONAGLGSNALSIGG-----GNQDVTYNQLAGL 60  
 Db 353 GCGGNMNMFGSGGAGGGGTGLAGTGVNFGSIANPTGANGTDSNGNGMOT----- 404  
 QY 61 TCGMMNMNMGGGGLMGGLGGLG--NGLGSGGLGEGLSNALNDMLGSLNTLGSKGG 118  
 Db 405 -----CGNG--GPGAGGVGERAGVGCGGGLGESLDG--NNGTGG-----KGG 443  
 QY 119 NNTTSTNSPLDQALGINSTQNDSTSGTSTSSDPMQQLKMFSEIMQSLFPGDOP 178  
 Db 444 AGGTACTDGCAGAGGAGGAGGIGEFHDSAGV-----ATGCGGD 481  
 QY 179 GTGGSSSGGKQPEE--QNAKKGYTDAL---SGLMNG--LSQLLNGGLG----- 225  
 Db 482 GATGVGDGCGAGGAGGCGGCHNTGVDAFEGDGGIGDNGALGAAGNGCTGAGAGNGG 541  
 QY 226 -----GCGMAGTGTGLSSLSGKGLQNLSPY---DYQQLGNAVGTGIGMKAGIALND 276  
 Db 542 RCGMLLNGAGAGAGGTGCT--GGGGAAGFRAGVGAGGEGELTGTAGTAGTGGTGLGGLG 600  
 QY 277 IG 278  
 Db 601 VG 602

## RESULT 13

hypothetical glycine-rich protein Rv0278c - Mycobacterium tuberculosis (strain H37RV)  
 C: Species: Mycobacterium tuberculosis  
 C: Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C: Accession: D70835  
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
 A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A: Reference number: A70500; MUID: 98295987  
 A: Accession: D70835  
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A: Molecule type: DNA  
 A: Residues: 1-957 <COIL>  
 A: Cross-references: GB:AL021930; GB:AL123456; NID:93261524; PIDN: CA17353.1; PID: 9290946  
 A: Experimental source: strain H37RV  
 C: Genetics:  
 A: Gene: RV0278C  
 C: Superfamily: elastin

Query Match 11.6%; Score 241.5; DB 2; Length 957;  
 Best Local Similarity 25.8%; Pred. No. 7.3e-08;

Matches 117; Conservative 42; Mismatches 149; Indels 145; Gaps 21;  
 QY 2 SLNTSLGASTMISTIGAGGNNGLGTSRONAGLGG-----NSALDGGGNQNDVTYNQLA 57  
 Db 362 TLTAGIGGA-----CGAGNAGLLFGSGSGAGGFGFPADGGGPGGN-----A 407  
 QY 58 GLTTGMMNMNMGGGGLMG--GGTGGGLGNGLGSGL-----GEGLSNAL---N 103  
 Db 408 G-----TVGSSGAGNGCGVCGGFPAGIGAGAGCTGGLGNGGNGGASAVTGCN 458  
 QY 104 DMLGSLNTLGSKGNNTTSTNSPLDQALGINSTQNDSTSGTSTSSDPMQQLK 163  
 Db 459 GGICTGTVLIGN--CGNGSGGGLGAGKAGVGSLLGLDGFNAFAPSTSPHTLQDNVNL 517  
 QY 164 MFSEIMOS-----LFGDGDGDTGGS-----SSGKQTEBGRQNAKKGYTDALSG--LMNG 213  
 Db 518 VNEPPTLGRPLDNGANGTGTGADGAGGWLFGNANGTPTGTGAAGAGGWLFGNG 577  
 QY 214 LSQLLNGGLGAG-----GCGNAGTGLDSSSLGKGLQNLSPDY 254  
 Db 578 -----CGGCHNTMTATNAGAGAGGILFPGCGGTG--GIATGAGIGAGAGAGV 630  
 QY 255 QQLGNAVGTG-----IGKAGIQALNDIGTNRHSSTFSFVNGKDRAMAKETIGQFMDOYR 308  
 Db 631 SLTIGSGGTGNGCNSIGV--AGIGAGAGRG-----GDAGL----- 664  
 QY 309 EVYGRKQYQKPGQPEVKTDDKSWAKLSRPDDGKTPAMQEPNKAKKIKRPMAGTCN 368  
 Db 665 -LFG-----AAGTGG-----HGAAGVPAGVGVGAGGN 690  
 QY 369 GNLQAGAGSSSLGIDAMAGDAINNMALGKLG 401  
 Db 691 GGLFANGAGAGAGFNA--AGCNGCNGGLCTGTG 722

## RESULT 14

hypothetical glycine-rich protein Rv1840c - Mycobacterium tuberculosis (strain H37RV)  
 C: Species: Mycobacterium tuberculosis  
 C: Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C: Accession: H70663  
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
 A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A: Reference number: A70500; MUID: 98295987  
 A: Accession: H70663  
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A: Molecule type: DNA  
 A: Residues: 1-515 <COIL>  
 A: Cross-references: GB:283859; GB:AL123456; NID:93261678; PIDN: CAB06114.1; PID: 917812  
 A: Experimental source: strain H37RV  
 C: Genetics:  
 A: Gene: Rv1840C  
 C: Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 11.5%; Score 238.5; DB 2; Length 515;  
 Best Local Similarity 27.3%; Pred. No. 5.5e-08;  
 Matches 109; Conservative 23; Mismatches 128; Indels 139; Gaps 15;

QY 18 GCGAGNNGLLGTSRONAGLGSNALSIGGNNQNDVTYNQLAGLLTGMNMNMNMGGGGLMG 77  
 Db 148 GCGAGGAGLIGNG--GNGGIGGPGATGLAG-----AGVGGLLFGDGGNGAGGGLG 197  
 QY 78 GGLGGGLGNGLGSGLGGLGSLNALNDMLGSLNTLGSKGNNTTSTNSPL--DQALGI 135  
 Db 198 TGPVGAATG--GIGGPGGAAGVLFEGHAGAGAGGAGKAGFAGAGAGTGTGGLLYGNGGNG 256  
 QY 136 NSTSQNDSTSGTSTSSDPMQQLKMFSEIMQSLFPGDGDCTGSSSGGKQPEBQ 195

Db 257 NPSGAADGAGADA-----RLINGGDG-----GSYGAAPTG--- 289  
QY 196 NAYKGYTDALSGLMGNSQLGNGSLG---GCGNAGTGLDSSSLGCKGLNLSGPV 252  
Db 290 -----IGN------GNGGNGMLYGDSGSGTLDGFSGSGTG----- 322  
QY 253 DYQOLGNAVGTGIGKACIQAALNDIGTRHSSTRSFVANKGPRAMAKEIGQFMDQYPEVFG 312  
Db 323 -----GNNGMFGDGNCGSFDFDNGSGDGTGCTILGNGD----- 358  
QY 313 KPQYQKPGQEVKTDKSWAKALSKPDDGMPASMEQFNKAKMKRPMAGDTGN--- 368  
Db 359 -----GNSVQTD-----GFL-RHGCGDGNVGL 382  
QY 369 -GNLQARGGSSIGIDAMMAGDAINN---MALCKLGA 402  
Db 383 IGNAGAGAGSNGTCVFAPGGSGGNGGNCALLVGNGGA 421

## RESULT 15

B70812  
hypothetical glycine-rich protein Rv0834c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: B70812  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987  
A:Accession: B70812  
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-882 <COL>  
A:Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PIDN:CAAI7640.1; PID:g291689  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: Rv0834c  
C:Superfamily: elastin

Query Match 11.5%; Score 238.5; DB 2; Length 882;  
Best Local Similarity 27.2%; Pred. No. 1e-07;  
Matches 113; Conservative 36; Mismatches 128; Indels 139; Gaps 23;

QY 7 GLGASTMIOISIGAGGNN-GLIGTSRONAGLGNALIGL-----GNNNDTVNQLAGIL 60  
Db 521 GGGASVGTGGSGAGGDFGVGA-----GNGGNMAGTGVGVANGCN-----GSSA 568  
QY 61 TGMAMMMAMGCGGLMGGGLGGLGNGLGGSGGLGSLNALN--DMLGSLNT-LSKRG 117  
Db 569 TG---ALAAVCGAGAGAGATSGTG-GFGAGAGSARGLIFALGAGAGAGDASTGVGCPG 624  
QY 118 GNNSTSTNSPLDQALGINST-SQNDSTSTGTDSTSDSSDPMQQLKMFSEIMQSLFPG 176  
Db 625 GFGGTGTASSPFGIALAIGAGAGAGATSGATGGAGDGVFEGT-----AVLGLDFG-G 678  
QY 177 QDGTGSSSGGKQPTGEGQNAKKGVTDLASL--MGNCLSQLG-----NGGLGGGCGN 230  
Db 679 AAGAGAGATG-----DGAITGAGGFGGAGAGIAFLGFSVYLHGAGGAGGTA 725  
QY 231 AGTGLDSSLGSKGLQNLGSPVYQOLGNAVGTGIMKAG-----IQALNDIGTRHSST 285  
Db 726 TGTGNGGAGCGGG---LSSPV-----ILGIGAGAGCGDGGALGVLCMGCG----- 769  
QY 286 RSFVKKGDAMAKEIGQFMDQYPEVGRKFOYQKPGQEVKTDKSWAKALSKPDDGMP 345  
Db 770 ----DGDGGEAVAVGIAG-----GAG----- 788  
QY 346 ASMEQFNKAKMKRPMAGDTGNLQARGAGSSSLGIDAMMAGDAINNMLCKLG 401

Db 789 -----GAGC-----AAPTGNG-----GAGCN-----GSDALGLVGVGNG 818

Search completed: October 25, 2001, 11:44:38  
Job time: 74 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 11:44:19 ; Search time 13.1 Seconds  
(without alignments)  
1053.814 Million cell updates/sec

Title: US-09-412-100-23

Sequence: 1 MSINISGLASTMTQISIGGA.....DAMAGDAINNALGKLGAA 403

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2050	98.6	403	1 HRPN_ERMAN	001099 erwina amy
2	828.5	39.9	356	1 HRPN_ERMCA	047279 erwina car
3	718.5	34.6	340	1 HRPN_ERMCH	047278 erwina chr
4	248	11.9	463	1 YAB6_MYCTU	053416 mycobacteri
5	242	11.6	778	1 Y034_MYCTU	P71933 mycobacteri
6	241.5	11.6	957	1 Y278_MYCTU	P56877 mycobacteri
7	238.5	11.5	515	1 Y140_MYCTU	050594 mycobacteri
8	238	11.4	1901	1 Y208_MYCTU	053553 mycobacteri
9	235.5	11.3	914	1 WA22_MYCTU	006794 mycobacteri
10	234.5	11.3	543	1 YP91_MYCTU	050630 mycobacteri
11	233.5	11.2	338	1 GRP_ARATH	P27483 arabidopsis
12	233.5	11.2	434	1 YK98_MYCTU	010707 mycobacteri
13	232	11.2	801	1 Y747_MYCTU	053810 mycobacteri
14	229.5	11.0	603	1 YD25_MYCTU	010637 mycobacteri
15	214.5	10.3	384	1 GRP1_PERTY	P09789 pecunia hyb
16	213.5	10.3	481	1 LORI_MOUSE	P18165 mus musculu
17	211.5	10.2	747	1 SPDI_NEPEL	P19837 nephila cla
18	210	10.1	1156	1 Y118_MYCTU	076743 caenorhabdi
19	202.5	9.7	498	1 Y118_MYCTU	050615 mycobacteri
20	199	9.6	937	1 HRY1_CANAL	P46591 candida alb
21	196.5	9.5	465	1 GRP2_PHAVU	P10496 phaseolus v
22	193	9.3	183	1 GRP2_ORISA	P23834 oryza sativ
23	191	9.2	252	1 GRP1_PHAVU	P10495 phaseolus v
24	191	9.2	316	1 LORI_HUMAN	P23490 homo sapien
25	190	9.1	641	1 EBNI_EBV	P03211 Epstein-Bar
26	186.5	9.0	1113	1 N116_YEAST	002630 saccharomyc
27	186	8.9	749	1 TROP_HUMAN	Q12816 homo sapien
28	184.5	8.9	419	1 CSP_PLACM	P08676 plasmodium
29	178.5	8.6	959	1 N100_YEAST	002629 saccharomyc
30	177.5	8.5	1224	1 PER_DROME	P07663 drosophila
31	174.5	8.4	672	1 PHX5_MOUSE	P08399 mus musculu
32	174	8.4	212	1 BEG1_SCHYA	P19470 schistosoma
33	174	8.4	700	1 NONA_DROME	Q04047 drosophila

34	173.5	8.3	1380	1 DDX9_MOUSE	070133 mus musculu
35	172	8.3	5263	1 FB0H_BOMMO	P05790 bombyx mori
36	171.5	8.2	1208	1 PER_DROMA	Q24767 drosophila
37	169	8.1	526	1 FUS_HUMAN	P35637 homo sapien
38	169	8.1	1596	1 MAM_DROME	P21519 drosophila
39	167	8.0	401	1 CSP_PLACG	P08674 plasmodium
40	166.5	8.0	393	1 CSP_PLACB	P14593 plasmodium
41	165.5	8.0	207	1 BGG2_SCHYA	P19469 schistosoma
42	164.5	7.9	404	1 CAZ_DROME	Q27294 drosophila
43	164	7.9	518	1 FUS_MOUSE	P35635 mus musculu
44	164	7.9	622	1 RICI_HUMAN	P35527 homo sapien
45	160	7.7	182	1 K2C3_BOVIN	P04261 bos taurus

## ALIGNMENTS

## RESULT 1

HRPN\_ERMAN STANDARD: PRT: 403 AA.

AC 001099;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE HARPIN (HARPIN-EA).

GN HRPN.  
OS Erwina amylovora.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Erwina.  
OX NCBI\_TaxID=552;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.  
RC STRAIN-EA321;  
RX MEDLINE=92320301; PubMed=1621099;  
RA Wei Z.-M., Laby R.J., Zumoff C.H., Bauer D.W., He S.Y., Collmer A.,  
RA Beer S.V.;  
RT "Harpin, elicitor of the hypersensitive response produced by the  
RT plant pathogen Erwina amylovora.";  
RL Science 257:85-86(1992).  
RN [2]

## REVIEWS.

RC STRAIN-EA321;  
RA Laby R.J., Kim J.F., Beer S.V.;

Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.

-1- FUNCTION: ELICITS THE HYPERSENSITIVE RESPONSE (HR) IN THE PLANT  
UPON INFECTION. HARPIN ELICITS HR IN NON-HOSTS AND IS ALSO

REQUIRED FOR PATHOGENICITY IN HOST PLANTS.

-1- SUBCELLULAR LOCATION: SECRETED; VIA THE HRP SECRETION PATHWAY.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; M92994; AAC31644.2; -  
CC DR Hypersensitive response.  
CC KM DOMAIN 1 270 GUY-RICH.  
CC FT DOMAIN 1 270 POLY-MET.  
CC SQ SEQUENCE 403 AA; 39697 MW; 146FA642351DBE87 CRC64;

Query Match 98.6%; Score 2050; DB 1; Length 403;

Best Local Similarity 99.0%; Pred. No. 4.1e-112;  
Matches 399; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	1	MSINISGLASTMTQISIGGAGNNGILGTSRONAGLGNSALGLGGGNDVYNQLAGLL 60
DB	1	MSINISGLASTMTQISIGGAGNNGILGTSRONAGLGNSALGLGGGNDVYNQLAGLL 60
QY	61	TGNMAMMSMAGGGGILMGGLGGGLGGGLGGSGGLGEGCLSNALNDMLGGSILNTLGSNGNN 120

D <b>b</b>	61	TGNNNNNMMGGCGSLMGGGLGCGAGNLDGSGCGELSNALNDMLGSLNTLGSKGNN	120
O <b>y</b>	121	TTSTTNSPLDQALGINSTSONDSTSGTDSTSDSSDPMOQLKMFSTIMOSLFEGDQDT	180
D <b>b</b>	121	TTSTTNSPLDQALGINSTSONDSTSGTDSTSDSPMOQLKMFSEIMOSLFEGDQDT	180
O <b>y</b>	181	OGSSSSGKOPESBOMNYKKGGVYDALSGLMGNGLSOLLNGNLGGGCGGAAGTGIDGSSL	240
D <b>b</b>	181	QGSSSSGGQPEGBONMYKKGVYDALSGLMGNGLSOLLNGNLGGGCGGAAGTGIDGSSL	240
O <b>y</b>	241	GGRKLQNLSGPVDYQOQAGNAVGTIGCKACIQIALINDIGTRHSSSTRFVAKGGRAMAKEI	300
D <b>b</b>	241	GGRKLQNLSGPVDYQOQAGNAVGTIGCKACIQIALINDIGTHSDSTRFVAKGGRAMAKEI	300
O <b>y</b>	301	GQPMDOYIEPVFGKPYOKYCPGEQVEKYTDKSMWAKLSPPDDGWTPTASMEQFNRAKGMIKR	360
D <b>b</b>	301	GQFMDOYPPEVFGKPYOKYCPGEQVEKYTDKSMWAKLSPPDDGWTPTASMEQFNRAKGMTKS	360
O <b>y</b>	361	PNAGDTCNGNLOARAGAGSSSLGIDPAMMAAGAINNMALGKTGAA	403
D <b>b</b>	361	AACGDTGNGNLOARAGAGSSSLGIDPAMMAAGAINNMALGKTGAA	403

RESULT	2
ID	HRPN_ERMCA
AC	047279;
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	15-DEC-1998 (Rel. 37, Last annotation update)
DE	HARPIN (HARPIN-ECC) (FRAGMENT).
GN	HRPN.
OS	Erwinia carotovora.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC	Pectobacterium.
OX	NCBI_TaxID=554;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=71;
RX	MEDLINE=96405946; PubMed=8810071;
RA	Cut Y., Madl L., Mukherjee A., Dumenyo C.K., Chatterjee A.K.;
RT	"The RsmA-mutants of Erwinia carotovora subsp. carotovora strain
RT	Ecc71 overexpress hrpNCC and elicit a hypersensitive reaction-like
RT	response in tobacco leaves.";
RL	Mol. Plant Microbe Interact. 9:565-573(1996).
CC	-1- FUNCTION: ELICITS THE HYPERSENSITIVE RESPONSE (HR) IN THE PLANT
CC	UPON INFECTION. HARPIN ELICITS HR IN NON-HOSTS AND IS ALSO
CC	REQUIRED FOR PATHOGENICITY IN HOST PLANTS.
CC	-1- SUBCELLULAR LOCATION: SECRETED; VIA THE HRP SECRETION PATHWAY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; L78834; AAB4973.1; -
KW	Hypersensitive response.
FT	DOMAIN 1 224
FT	NON_TER 356 356
FT	SEQUENCE 356 AA; 35621 MW; 108BA6B9D27F9DE4 CRC64;
FT	GLY-RICH.

Query Match	39.9%	Score 828.5	DB 1	Length 356
Best Local Similarity	48.2%	Pred. No. 1.4e-41		
Matches 196	Conservative 37	Mismatches 115	Indels 59	Gaps 11

OY 3 LNTSGAGASTWQISIGGAGGNNGIL-GTSHONAGLGCNSALGLGGNONDTVNOLAGLLT 61  
|| : ||| :||: ||| | | | | | :||: ||: ||:  
Db 2 LNSLGCGAS-LQITI-KAGGNGLEPSQSSONGSGSPQSAF---GGQRNIAEQLSIDIMT 56

Oy 62 GMAHMAHMAAGGGLMGGLG-----GGLGGLGSSGLEEGSLNADMLNDLGGSLNMTGS 115  
 Db 57 TMTFPGSMAGG--MSGGLGGLSSLSGLGGLG--GLGGLGGLSSLSGSLGSLG--GG 111  
 Oy 116 KGNNTTSTNSPLDQALGINSTSQNDSTGTSTSDSSDPWQILKMFSEIMQSLFGD 175  
 Db 112 LGG-----ALGAGMNAHNSPAMMS-----LTFSALEDLGLG 143  
 Oy 176 GODGTGSSSGGKOPTEBEONAYKKGYTDLALSGMLNGSLGLGNGGLGGGGGAGTGL 235  
 Db 144 GMSQGGGLFPGKOPSPSEISATYQGVADLSALGLNGLSQTKQ-----T 189  
 Oy 236 DGSSLGGGLGLOLSPVDYQOLGANAAGVGTGCMKAGIQALNDIGTHRSSSTPSFYNKGDRA 295  
 Db 190 SPLQIGNNGLGLGSAGAFNLOGLSTLGSSTVQKGLGLOELNNISTHNDSPRYFYDKEDRG 249  
 Oy 296 MAKEIGEMPDQYPEVEFGKFPQYQKSPGQEVKTKDDSPMAKALSKPPDDGMPFASMOFKAK 355  
 Db 250 MAKEIGEMPDQYPEVEFGKAEYQKDNWQYAKQEDKSPMAKALSKPPDDGMPGSMQKFKKAV 309  
 Oy 356 GMIKRPMAQDPTGNGLQARAGGSSSLGTDAMMADDAIINNALGKTLGA 402  
 Db 310 GMIKSAIRGDTGNTLNSARGGASLSLGDIAAMIDRIYVNGLTKLSS 356

ID	HRPN_ERMCH	STANDARD:	PRT:	340 AA.
AC	047278;			
AD	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	HARPIN (HARPIN-ECH).			
CN	HRPN.			
OS	Erythrina chrysanthemi.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Pectobacterium.			
OX	NCBI_TaxID=556;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-EC16 / AC1450;			
RX	MEDLINE=96172740; PubMed=8589405;			
RA	Bauer D.W., Wei Z.M., Beer S.V., Collmer A.;			
RT	"Erythrina chrysanthemi harpinEch: an elicitor of the hypersensitive			
RT	response that contributes to soft-rot pathogenesis.";			
RL	Mol. Plant Microbe Interact. 8:484-491(1995).			
CC	-1- FUNCTION: ELICITS THE HYPERSENSITIVE RESPONSE (HR) IN THE PLANT			
CC	UPON INFECTION. HARPIN ELICITS HR IN NON-HOSTS AND IS ALSO			
CC	REQUIRED FOR PATHOGENICITY IN HOST PLANTS.			
CC	-1- SUBCELLULAR LOCATION: SECRETED; VIA THE HRP SECRETION PATHWAY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; L39897; AAC31978.1; .			
KW	Hypersensitive response.			
FT	DOMAIN 1 203			
SEQUENCE	340 AA: 34274 MW; EAD82C23731E4F4C5 CRC64;			
GLY-RICH				

Query Match	34.6%;	Score 718.5;	DB 1;	Length 340;
Best Local Similarity	42.6%;	Pred. No. 3.1e-35;		
Matches 1/2; Conservative	41;	Mismatches 114;	Indels 77;	Gaps 10;

```
Oy      13 MOISI-GGAGGNNGLGTSRONAGLGG-NSA--LGLGGNQNDVTNQLAGLLTGMMMM 6778
        |||:|   ||: |: |   ||| |||   ||   :||::|   ||| ||
Db      1 MOTTKAHIGDGLVSGLGLGAAGLKLGNLSAASLSSVDKLTSTIDKLTSLATSM--- 5779
```

QY 68 SMWGGGGLMGGLGGLGNGELGEGLSNALNDMLGSLNTLGSKGNNTTSTNS 127  
 DB 58 -----FGALAGGLGAS-SKGLGMSKOLGOSFEN-----GAGASNLISVPK- 98  
 QY 128 PLDQALGINSTSONDSTSGTSDTSDSDPMOOLKMFSEIMOSLEF-----DQDQD 179  
 DB 99 -----SGDALS-----KMFDAALDDLGHDTVTKLTNNSQ 130  
 QY 180 TQSSSSGKOPTEGEONAYKKVYTDALSLGLMGSLQLLNGGLGGGCGNAGTGLDSS 239  
 DB 131 LANSMLNASQMTQGNNAFESGVNNAISLILGNGLGOSM-----SGFSQPS 176  
 QY 240 LGGAGLONLSGPVYQOLGNAVGTGIGMKAGLQALNDIGTHRSSPTFSFYKNGDRMAKE 299  
 DB 177 LGGAGLQGLSGAGAFNOLGNAIGVGQNALSLNSVSTHVDGNHHEVDKEDRGMAKE 236  
 QY 300 IGFMDQYEPVEFGKPOYOKGPGQEVKTDDSKSMKALSKPDDCGMTFPAEMQFNKAKMIK 359  
 DB 237 IGFMDQYEPVEFGKPOYOKGPGQEVKTDDSKSMKALSKPDDCGMTFPAEMQFNKAKMIK 296  
 QY 360 RPAAGTGNGNTLQARGAGSSILGIDAMAGDAINNALGKLGAA 403  
 DB 297 SAVAGDTGNTNMLRGAGASLGDIAAVYGDKIANNISLGKLANA 340

RESULT 4  
 YAG8\_MYCTU STANDARD; PRT; 463 AA.  
 ID YAG8\_MYCTU  
 AC 053416;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE HYPOHETICAL PE-PGRS FAMILY PROTEIN RVL068C.  
 GN RVL068C OR MTV017.21C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS  
 SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AL021897; CAB17184.1; -  
 DR Tuberculist: Rv1068c; -  
 DR InterPro: IPR000084; -  
 DR InterPro: IPR002952; -  
 DR Pfam: PF00934; PE; 1.  
 DR PRINTS: PRO1228; EGGSHLL.  
 KM Hypothetical protein.  
 SO SEQUENCE 463 AA; 39305 MW; CF5696A7E9593952 CRC64;

Query Match 11.98; Score 248; DB 1; Length 463;  
 Best Local Similarity 31.88; Pred. No. 6,7e-08;  
 Matches 93; Conservative 17; Mismatches 112; Indels 70; Gaps 13;

QY 8 LGASTMOISIGAGGNGGLGTSTRONAGLSNALSILGSGNODTYVQLGLLTGMMMM 67  
 DB 203 IGAPVAGAGAGAGAGTAGLGE-----NGAGAGAGAGAGAGRGD--GSGAGMLSGNGDA 256  
 QY 68 SMWGGGGLMGGLGGLGNGELGEGLSNALNDMLGSLNTLGSKGNNTTSTNS 127  
 DB 257 GTGGGGGNAGGNGGAGAGMLSGNGGTGGG-----GGTAGA-GGGGNG-----NS 301  
 QY 128 PLDQALGINSTSONDSTSGTSDTSDSDPMOOLKMFSEIMOSLEFDDGDDTQSS--- 184  
 DB 302 GIDPG---NGOGADTGNAGCGHGSA-----AKLFGDGGAGAGAGMGST 344  
 QY 185 -----SGKQPTGEONAYKKVYTDALSLGLMGSLQLL-----GNGGLGGGQ--G 229  
 DB 345 GGTGGGGGFGGTTGGNGNGHAGAGSGGTAGLSSGSGGTGGDGGNGGLGSGAKG 404  
 QY 230 NAGTGLDSSISGKGLONLSGPVYQOLGNAVGTGIGMKAG---TQALNDIG 278  
 DB 405 NGNGGD-----GKG-----GDAQLINGNGNGNGKGTGLMPGINTG 445

RESULT 5  
 YQ34\_MYCTU STANDARD; PRT; 778 AA.  
 ID YQ34\_MYCTU  
 AC P71933;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE HYPOHETICAL PE-PGRS FAMILY PROTEIN RV2634C.  
 GN RV2634C OR MTCY441.04C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS  
 SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: Z80225; CAB02341.1; -  
 DR HSSP: P04002; IATF.  
 DR Tuberculist: Rv2634c; -  
 DR InterPro: IPR000084; -  
 DR InterPro: IPR001899; -  
 DR Pfam: PF00934; PE; 1.  
 KM Hypothetical protein.

SEQUENCE 778 AA: 63131 MW: DAB20FE58E4999E7 CRC64:

Query Match 11.6%; Score 242; DB 1; Length 778;  
Best Local Similarity 30.1%; Pred. No. 2.6e-07;  
Matches 91; Conservative 19; Mismatches 110; Indels 82; Gaps 14;

Y 7 GLGASTMOISIGAG--GNNGLGTSRONAGLGSALGLAG-----GNQNDTVNOLAGLL 60  
Db 353 GNGGNANMFSGAGGCGCTGAGTNGVNPISIANPNTGANCNDNSNGNQ----- 404  
Y 61 TGMNMMNMGCGGLMGGLGGLG--NCLGSSGGLGEGLSNMLDMLGSLNTLGSKGG 118  
Db 405 -----GNG--GGPAGVGEAGVGAGGGLGEGSLDG--MDGTGG-----KGG 443  
Y 119 NNTTSTNSPLDQALGINSTSQNDSTSGTSDTSDSPMQLLKMFSEIMQSLFGDGD 178  
Db 444 AGGTAGTGGAGAGAGAGAGIGTDSAGV-----ATGGEAGD 481  
Y 179 GTGSSSGGKQPTGEG--ONAYKKGYTDAL---SGLMNG---LSQLLNGGLAG----- 225  
Db 482 GATGVDGVDGAGGAGGAGGHTGVDAGFGGCGIGGDNALGAAGNGGTGAGAGNGG 541  
Y 226 -----GGGNGAGTGLDSSLGKGLONISGPV---DYQLGNANGTGTGMAAGTOLAND 276  
Db 542 RCGMLIGNGAGAGAGTGT--GGGAGAGFAGVGAGGEGTLVDAGTGTGGLGGLAG 600  
Y 277 IG 278  
Db 601 VG 602

RESULT 6  
Y278\_MYCTU STANDARD: PRT: 957 AA.  
AC P56877:  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE HYPOTHEITICAL PE-PGRS FAMILY PROTEIN RV0278C PRECURSOR.  
GN RV0278C OR MTW035.06C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID-1773;  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV:  
RX MEDLINE-98295987; PubMed-9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Horsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter K., Seeger K., Skelton S., Squares S., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrell B.G.;  
RT Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
-I- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS  
SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch))  
CC  
CC EMBL: AL021930; CAA17353.1; -  
CC  
CC TubercuList: RV0278C; -

DR InterPro: IPR000084; -  
DR Pfam: PF00934; PE: 1.  
RN Hypothetical protein; Repeat; Signal.  
FT SIGNAL 1 30 POTENTIAL.  
FT CHAIN 31 957 HYPOTHEITICAL PE-PGRS FAMILY PROTEIN  
FT RV0278C.  
SQ SEQUENCE 957 AA: 81905 MW: 71EBAD417FBA47C CRC64:

Query Match 11.6%; Score 241.5; DB 1; Length 957;  
Best Local Similarity 25.8%; Pred. No. 3.4e-07;  
Matches 117; Conservative 42; Mismatches 149; Indels 145; Gaps 21;

Y 2 SLNTSLGASTMOISIGAGNNGLGTSRONAGLAG-----NSALGLGNGNDTVNOLA 57  
Db 362 TLTAGGIGA-----GGAGNAGLLFGSGCGAGAGGCFADGGGGRGN-----A 407  
Y 58 GLITGMMNMMNMGCGGLMG--GGLGGLGNGLGSGL-----GGLSNL--N 103  
Db 408 G-----TFVGSAGAGNGCGVGFAGGIGAGGTFGLGNGNGNGGASAVTGN 458  
Y 104 DMLGSLNTLGSKGNNTTSTNSPLDQALGINSTSQNDSTSGTSDTSDSPMQLLK 163  
Db 459 GGIGGTGVLIGN--GGNGSGGIGAGKAGVGVSLLLGLDGFNAPASTSPHTLQNNVLN 517  
Y 164 MFSEIMQS-----LFGDQDGTGGS---SSGKQPTGEGONAYKKGYTDALSG--LMNG 213  
Db 518 VYNEPFOITLITGRPLIGNANGTPTGTGADGAGGWLFGNANGTPTGTGAAGAGWLFNG 577  
Y 214 LSQLLNGGLG-----GCGNAGTGLDSSLGKGLONISGPV 254  
Db 578 -----GNGGHGATNTAATATGAGAGAGILFCTGNGGTG--GIATGAGIGAGAGAGV 630  
Y 255 QOLGNAYGT-----IGKKAGTOLNDIGTFRHSSTRFVKGRRAKELIGTMDQYR 308  
Db 631 SLILSGGTGTGNGNSIGV--AGIGAGGGR-----GAGL----- 664  
Y 309 EYFGKPYQYKPGQEVKTDKSMARALSKPDDGWTSPASMEQFNFKAKIKRPMAGDTGN 368  
Db 665 -LFG-----AAGTGG-----HGAAGVPAVGAGAGGN 690  
Y 369 GMLQARAGCGSSSLGIDAMAGDAINNALGKLG 401  
Db 691 GGLFANGAGAGAGFMA--AGNGNGGLFGTGG 722

RESULT 7  
Y140\_MYCTU STANDARD: PRT: 515 AA.  
ID Y140\_MYCTU  
AC Q50594;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DE HYPOTHEITICAL PE-PGRS FAMILY PROTEIN RV1840C.  
GN RV1840C OR MTCTY1A1.04 OR MTCTY359.33.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID-1773;  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV:  
RX MEDLINE-98295987; PubMed-9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Horsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter K., Seeger K., Skelton S., Squares S., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrell B.G.;  
RT Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";

RL Nature 393:537-544(1998).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGNS  
 CC SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z83859; CAB06114.1; -.  
 DR Tuberculin; RV1840C; -.  
 DR InterPro: IPR000084; PE; 1.  
 DR Pfam: PF00934; PE; 1.  
 KM Hypothetical protein; Transmembrane.  
 FT TRANSMEM 1 21 POTENTIAL.  
 FT TRANSMEM 165 185 POTENTIAL.  
 FT TRANSMEM 199 219 POTENTIAL.  
 FT SEQUENCE 515 AA; 43916 MW; 2EDC8B6C6D28B7E3 CRC64;  
 SQ  
 Query Match 11.5%; Score 238.5; DB 1; Length 515;  
 Best Local Similarity 27.3%; Pred. No. 2.7e-07;  
 Matches 109; Conservative 23; Mismatches 128; Indels 139; Gaps 15;  
 QY 18 GGAGGNNGLGTSRQNGALGNSALGGLGQNDVTNOLAGLTTGMMAMSMGSGGLMG 77  
 DB 148 GGAGGADAGLIGNG-GNGSIGPGATGLAG-----AGVGGLLKGDDGNGAGGLG 197  
 QY 78 GGLGGGGLNGLGGSGGLGGLSNALNDMLGSLTLTSGKGNNTTSTNSPL--DQALGI 135  
 DB 198 TGPVGATG-GTGGPGGAAGVGLFGHGGAGGAGGLKAGPAGAGTGTGGLTYGNGNGG 256  
 QY 136 NSTSNDSTGTSTSTSSDPMDQLKMFSEIMQSLFGDDODTQSSSGSKOTBEBO 195  
 DB 257 NVPSGADGAGGDA-----RLIGNGDG---GSVCAATPG--- 289  
 QY 196 NAYKKVTDALSGLMGSLGGLGGLG---GCGNAGTGLDSSSLGSKGLQMLSPV 252  
 DB 290 -----IGNG-----GNGNGMLHYGDCSGSLTQFSOQGTG----- 322  
 QY 253 DYQOLGNAVGTGIGKACIGIALNDIGTRHHSSTRSFVKKGDRAAKETIGQFMDOYPEVFG 312  
 DB 323 -----GNMGMGDDGNGGFFFDGNGDGGTGLIGNGD----- 358  
 QY 313 KPOYOKGPGOEKTKDDKMAKALSRPDDGMPASMEQFNKAKMIKPMAGDTGN--- 368  
 DB 359 -----GNSVQTD-----GFL-KHGHDGDNNAVGL 382  
 QY 369 -GNLQARGAGSSSLGIDAMAGDAIIN---MALKTLGA 402  
 DB 383 ICGNGAGAGAGAGTGVPAFGSGSGNGGNGALLVNGCA 421  
 RESULT 8  
 Y208 MYCTU STANDARD; PRT: 1901 AA.  
 AC 053553;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE HYPOTHEICAL PE-PGNS FAMILY PROTEIN RV3508 PRECURSOR.  
 GN RV3508 OR MTW023.15.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;

RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hounsby T., Jagels K., Kiroh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RA complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGNS  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AL022022; CAAL1745.1; -.  
 DR HSSP: P19972; IKVD.  
 DR Tuberculin; RV3508; -.  
 DR InterPro: IPR000084; -.  
 DR Pfam: PF00934; PE; 1.  
 KM Hypothetical protein; Repeat; Signal.  
 FT SIGNAL 1 30 POTENTIAL.  
 FT CHAIN 31 1901 HYPOTHEICAL PE-PGNS FAMILY PROTEIN  
 FT SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;  
 SQ  
 Query Match 11.4%; Score 238; DB 1; Length 1901;  
 Best Local Similarity 27.6%; Pred. No. 1.1e-06;  
 Matches 124; Conservative 29; Mismatches 180; Indels 116; Gaps 23;  
 QY 9 GASTMOISI-GGAGGNNGLGTSRQNGALGNSALGGLGQNDVTNOLAGLTTGMMAMM 67  
 DB 1133 GASTTISINANGAGAGNGGICG---KGGAGGAGTGLVGGSGGTGDDGAG----- 1179  
 QY 68 SMNGGGLMGGLG---GGLGN-GLGGSGGLG-EGISNALNDMLGSLTLTSGKGNNTTS 123  
 DB 1180 --SGGGGGGFGAAGRAAGGGGNGGVDGDEGASGLGLISGFDGQ---GGGAGAGSA 1233  
 QY 124 TTNSPLDQALGINSTSQNDSTSGDSTSDSSDPMDQLKMFSEIMQSLFG--DGODGTQ 181  
 DB 1234 -----GAGGINGAG---GAGGTGGAGCGCAP-----ATLIGPDDGDDGQ 1270  
 QY 182 GSSSGGKQPTGEONAYKKGVY---DALSGLMGSLGGLGGLG---GGGAGNAGTGL 235  
 DB 1271 GGIGG-----DGNMGFGAGVPGDDGNGAGFAGVP---GDGIGTGAGAGAGAGA 1322  
 QY 236 DG-----SSLGKGLNLGSPVDYQ-----LGNA 260  
 DB 1323 DCDPSIDGGGAGAGGCHGGGCGGLNSTGLASAAGDGGNGAGAGAGCGDDEFTIGS 1382  
 QY 261 VGT-GTGMKAGIOAL-NDIGTRHHSSTRSFVKKGDRAAKETIGQFMDOYPEVFGKPOYK 318  
 DB 1383 GGTGGTGDRAGVGGGLANTGTCGNMGITGAGRGDGGGAGSALSDGNGFRAGGGGOG 1442  
 QY 319 GPQOEKTKDDKMAKALSRPDDGMPASMEQFNKAKMIKPMMA---GDTGNGNLQAR 374  
 DB 1443 GVGNGNAGAGINGAGGTG---GTGAGAGDQNGNTTVAASGAGAGGGGDDGGGGI--- 1494  
 QY 375 GAGGSSLGIDAMAGDAIINMALGKLGAA 403  
 DB 1495 GGAGGNAAGFAGVPGD---GIGGTGGA 1519  
 RESULT 9



WA22	MYCTU		STANDARD;	PFR;	914 AA.
ID	MA22_MYCTU				
AC	006794:				
DT	30-MAY-2000	(Rel. 39,	Created)		
DT	30-MAY-2000	(Rel. 39,	Last sequence update)		
DT	30-MAY-2000	(Rel. 39,	Last annotation update)		
DE	WAG22 ANTIGEN PRECURSOR.				
G8	WAG22 OR RV1759C OR MYC28-25C.				
GN	Mycobacterium tuberculosis.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1773;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-H37RV;				
RX	MEDLINE=98295987; Pubmed=9634230;				
RA	Cole S.T., Brosch R., Parkhill J., Garner T.E., Churcher C., Harris D.,				
RA	Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F.,				
RA	Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Davies R., Devlin K., Felwell T., Gentile S., Hamlin N., Holroyd S.,				
RA	Hornisy T., Jagals K., Krogh A., Mclean J., Moule S., Murphy L.,				
RA	Oliver S., Osborne J., Quail M.A., Rajadream M.A., Rogers J.,				
RA	Rutser S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,				
RA	Taylor K., Whithead S., Barrell B.G.,				
RT	"Deciphering the biology of Mycobacterium tuberculosis from the				
RT	complete genome sequence."				
RL	Nature 393:537-544(1998).				
CC	-I- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS				
CC	SUBFAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> )				
CC	-----				
DR	EMBL; Z95890; CAB09322.1; ..				
DR	HSSP; P41140; ZSPA.				
DR	Tuberculist; RV1759c; ..				
DR	InterPro; IPR000084; ..				
KM	Plant; PF00934; Pf; 1.				
KM	Antigen; Repeat; Signal.				
FT	SIGNAL	1	30		POTENTIAL.
FT	CHAIN	31	914		WAG22 ANTIGEN.
QO	SEQUENCE	914 AA;	74354 MW;	F69553C3DBE866A8 CRC64;	

Query Match	11.3%	Score 235.5	DB 1	Length 914
Best Local Similarity	25.8%	Pred. No. 7,2e-07		
Matches	120	Conservative	38	Mismatches 173; Indels 135; Gaps 20;

  

QY	7	GLGASTWQISIGAGACGNNC----	LIGTSRONAGLCNSAL-----	GLGSGNQDVTYVQ	55
Db	186	GAGGSAMLFGAGAGAGAGCAATSLV	GGTIGCTGCTGAGNMLAGAACAGAGCFSESTAGG		245
QY	56	LAGL-LTGCMNNMSHMGCGGLMG-----	GGGCGGAC-NGIGASSGIGE----	GLSNA	101
Db	246	ACGAGAGAGGLFTTGGVYGAGCGG	GHTCGAGAGAGGLFAGAGMGAGCGFDHGTLTGTGA		305
QY	102	LNDMLGSLNTLGSKFGKNNTTSTTNS	PLDQALGINSTSONDDSTGTDSTSDSSDPMOOL		161
Db	306	GDDGGGGGLFCGAGGCGAGAGSG	GLTTGTGA-ACNGGNAAGTSLILAAAGAGAGGTGAGC----		359
QY	162	LKMFEIIMGSLPGCDODDPTGOS-----	SSGCGQPTPEGQENAVKKGTALSLGIMCG		213
Db	360	-----TYFGGKCGAGAGGAGN	AMLFSGSGGGGTG-----	FGFAAGAGGGGGS	404
QY	214	LSOLLGNGGLGAGGCG-----	GNAGTGLDGSISLGKGLONLSGP		251
Db	405	ACMLSGSGSGGAGSGGAPAGTAA	GAGAGAGAPGLTICNGMGNGGEGSGGTGAGVGAG-463		

```

0Y 2Y2 VDYQOLGNAVSTGJGKMRGIDA-----LNDICTHHSSSTRSVKNGDPAMA 297
Db 464 -----GNAVLIGNGEGEGIGALAGSGFGGGLLIGADDPYNPESTSPHN-----LQ 512
0Y 298 KEIQGFMDQYEFV--GKPYQKPGGGEKTKDDKMAKALSKPDDGNTPASMDFPKAKG 356
Db 513 QDILSEFNEPTEALTGRLIGNG-----DSGTPTGDD--GGAGG 550
0Y 357 MIKRPMAGDTCNGNLQARGAGSGSLGIDAMWAGAIINNMAIGKLG 402
Db 551 W----LFENGNGGAGAGTNGSGAG---AGGA--GGLFTGTGA 586

```

	RESULT	10
XP91_MYCTU		
ID	XP91_MYCTU	STANDARD; PRT; 543 AA.
AC	Q50630;	
DT	01-NOV-1997 (Rel. 35, Created)	
DT	01-NOV-1997 (Rel. 35, Last sequence update)	
DT	30-MAY-2000 (Rel. 39, Last annotation update)	
DE	HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV2591.	
GN	RV2591 OR MTCY227.10C.	
OS	Mycobacterium tuberculosis.	
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.	
OX	NCBI_TaxID=1773;	
RN	[1]	
RP	SEQUENCE FROM N.A.	

RX MEDLINE-98295987: PubMed-9634230:.  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gass S., Barry C.E. III, Tekela F.,  
 RA Baccroft K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Ratandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Sgares S., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.:  
 RT "Deciphering the biology of *Mycobacterium tuberculosis* from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PERS  
 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z77724; CAB01283.1; -  
DR Tuberculin; RV2591; -  
DR InterPro: IPR000084; -  
DR Pfam: PF00934; PE: 1.  
KW Hypothetical protein.  
SQ SEQUENCE 543 AA; 46287 MW; 59730339E5DD2DF59 CRC64;

```

Query Match 11.3% Score 234.5; DB 1; Length 543;
Best Local Similarity 28.0%; Pred. No. 4,8e-07;
Matches 111; Conservative 31; Mismatches 121; Indels 133; Gaps 21;

QY 18 GGAGGNGGILGTRONAGIGGNSALGIGGNGNDPYNLAGLLGTGMMMSMGG----- 72
      ||||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 213 GGAGGDAFLPIGCG-GNGGCGGFPAAGCGAGGN-----GCAGSGLEFVGAGAGVGGSS 264

QY 73 ---GGLMG-GGIGGIGKNGIGAGSGGIGEGLSN-----ALNDMLGSSLNTLGGKGNNTT 122
      ---|||---|||---|||---|||---|||---|||---|||---|||---|||
Db 265 EDVGGTGGAGGGAGGRFLGLGIGDGGAG-CTSNNGNGDGGAGGTAGGRFLSLGDDGNGCA 323

QY 123 STTNSPLDQALGINSTSDSTSGTGSTSDSSDPQQQLLKMFSEIIMOSLFFDGDGDDTGG 182

```



DB 324 GT-----AIGSNA-----GDGAGGDS-----ALIGAAGGSGG 353  
 OY 183 SSSGKQPTGEQNAAYKKVTADALSGLMGNGLSQLLGNGLGAGGQ---GGNAGTGIDGSS 239  
 DB 354 --LGG---FGESTGDDGGLGAGAVLIGTVG---GFGGLGGSSNCTGGAGAGGTGAT 404  
 OY 240 L-----GKGGLONLSGPVDYDQLGNAVGTGIGMKAGI-OALNDIGTHRSSTRSFVNKG 292  
 DB 405 LIGLGGAGGGGIGGGAFA-----VWVNGVVG-GLGGGGGGAALIGLAG-----GAG 449  
 OY 293 DRAMKEIGPFMDQYEPVFKPKYQKPGQGVETDOKSMALKSPDDGDMTPASMRQFN 352  
 DB 450 GAGCATVVG-----LCGNGGDX-----G 467  
 OY 353 KAKGMIRPMACD---TGNGNLQARGAGSSSLGIDA 385  
 DB 468 DGGGLFSTIGVGDDGNGNANGMANGNGNGNACVIA 503  
 RESULT 11  
 GRP ARATH STANDARD: PRT: 338 AA.  
 AC P27483:  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 01-AUG-1992 (Rel. 23, Last annotation update)  
 DE GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN PRECURSOR.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsids.  
 NCBI\_TaxID=3702;  
 RX MEDLINE=92003708; PubMed=1912511;  
 RA Ouligley F., Villiot M.L., Maché R.;  
 RT Nucleotide sequence and expression of a novel glycine-rich protein  
 RL Plant Mol. Biol. 17:949-952(1991).  
 CC -1- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: CELL WALL (POTENTIAL).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X58338; CAA41249.1; -  
 DR PIR: S17732; KMMU  
 DR HSSP: P30129; ADPV.  
 KW Cell wall; Structural protein; Repeat; Signal.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 338 GLYCINE-RICH CELL WALL STRUCTURAL  
 FT DOMAIN 21 338 GLY-RICH.  
 FT SEQUENCE 338 AA; 23691 MW; 046A6E8C1A4E89EB CRC64;  
 SO

Query Match 11.2%; Score 233.5; DB 1; Length 338;  
 Best Local Similarity 29.9%; Pred. No. 3.3e-07;  
 Matches 80; Conservative 14; Mismatches 77; Indels 97; Gaps 10;

OY 7 GLGASTMOISTGAGGNNGLGTSRONA--GIGNSALGIGGQNDTVNDLAGLLTGM 64  
 DB 59 GLG-----GGAGGGGIGGAGGAGGAGGAGGAGGGLGGH-----GGGIGG-- 100  
 OY 65 MMMSMGGGGLGAGGLGGLGNGLAG--SGGLGEGLSNADMDMLGSLNTYLGSKGNNTT 122

DB 101 -----GAGGAGGGLGCGHGGGIGGAGGSGGGLGGIGGAGGAGGGLGGH-- 152  
 OY 123 STNNSPLDQALGINSTSNDSTSGTSDSPMOQLKMFSEIMOSLFGDQDGTG 182  
 DB 153 -----GGGIGGAG 161  
 OY 183 SSSGKQPTGEQNAAYKKVTADALSGLMGNGLSQLLGNGLGAGGQGNAGTGIDGSSLGG 242  
 DB 162 GGAG-----GLGGHGGGIGGIGGAGGSGGGL-----GGGIGGAGGAGGAGG--GGAGG 209  
 OY 243 KGLONLSGPVDYDQLGNAVGTGIGMKAG 270  
 DB 210 GGLGGHG-----GFGGAGGGLGGAG 233

RESULT 12  
 YK98\_MYCTU STANDARD: PRT: 434 AA.  
 AC Q10707;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE HYPOTHEICAL 36.5 KDA GLYCINE-RICH PROTEIN RV2098C.  
 GN RV2098C OR MYCY49.38C  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 NCBI\_TaxID=1773;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hovnsby T., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skellern S., Squares S., Squires R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGNS  
 CC SUBFAMILY.  
 CC -1- CAUTION: THIS PROTEIN IS FRAMESHIFTED NEAR THE N-TERMINUS.  
 CC THE SEQUENCE HAS BEEN CHECKED BY AUTHORS IN REF.1 AND THEY REPORT  
 CC THAT NO ERRORS HAVE BEEN FOUND.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z73966; CAA98228.1; -  
 DR HSSP: P19972; IKVD.  
 DR Tuberculist; RV2098C;  
 DR InterPro; IPR000084; -  
 DR InterPro; IPR002952; -  
 DR Pfam; PF00934; PE; 1.  
 DR PRINTS; PRO1228; EGGSHELL.  
 KW Hypothetical protein.  
 SO SEQUENCE 434 AA; 36530 MW; B736E82CD1E0E54 CRC64;

Query Match 11.2%; Score 233.5; DB 1; Length 434;  
 Best Local Similarity 28.8%; Pred. No. 4.3e-07;  
 Matches 105; Conservative 22; Mismatches 150; Indels 87; Gaps 16;

OY 9 GASTMOISTGAGGNNGLGTSRONAGLGGNSALGIGGQNDTVNDLAGLL----- 60





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 11:43:59 ; Search time 24.18 Seconds  
(without alignments)  
2205.084 Million cell updates/sec

Title: US-09-412-100-23  
Perfect score: 2079  
Sequence: 1 MSLMTSGAGSTMQISTIGA.....DAMMAGDAINNALGKLGAA 403

Scoring table: BLOSUM62  
Gapop 10.0 / Gapext 0.5

Searched: 425026 seqs, 132305027 residues  
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP:REMBL.16:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mammal:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.podent:\*  
12: sp.unclassified:\*  
13: sp.vertebrate:\*  
14: sp.virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1181.5	56.8	382	2	Q9FCY8	Q9FCY8 erwina ste
2	1002.5	48.2	365	2	Q9KH45	Q9KH45 erwina her
3	736.5	35.4	344	2	Q9EXPO	Q9EXPO erwina chr
4	712	34.2	142	2	Q9LAW6	Q9LAW6 erwina amy
5	264.5	12.7	1489	2	Q53559	Q53559 mycobacteri
6	255	12.3	591	2	007224	007224 mycobacteri
7	249	12.0	731	2	Q50415	Q50415 mycobacteri
8	248	11.9	1381	2	Q53552	Q53552 mycobacteri
9	247	11.9	588	2	Q50396	Q50396 mycobacteri
10	246.5	11.9	1079	2	Q53557	Q53557 mycobacteri
11	246	11.8	1538	2	Q53395	Q53395 mycobacteri
12	245.5	11.8	1660	2	Q53215	Q53215 mycobacteri
13	245	11.8	923	2	Q53890	Q53890 mycobacteri
14	238.5	11.5	882	2	Q53845	Q53845 mycobacteri
15	236.5	11.4	714	2	Q53556	Q53556 mycobacteri
16	236.5	11.4	853	2	Q53439	Q53439 mycobacteri
17	235	11.3	439	2	Q06287	Q06287 mycobacteri
18	233.5	11.2	349	10	Q9LSP2	Q9LSP2 arabidopsi
19	233.5	11.2	615	2	Q05806	Q05806 mycobacteri

20	232	11.2	837	2	053684	053684 mycobacteri
21	229.5	11.0	1329	2	006810	006810 mycobacteri
22	228.5	11.0	667	2	053415	053415 mycobacteri
23	228.5	11.0	904	5	076271	076271 mytilus edu
24	228	11.0	783	2	053809	053809 mycobacteri
25	227	10.9	496	2	053631	053631 mycobacteri
26	226.5	10.9	491	2	006818	006818 mycobacteri
27	226.5	10.9	576	2	P71664	P71664 mycobacteri
28	226	10.9	694	2	053212	053212 mycobacteri
29	225.5	10.8	1011	2	P94986	P94986 mycobacteri
30	225	10.8	749	2	053844	053844 mycobacteri
31	224	10.8	1306	2	053775	053775 mycobacteri
32	223	10.7	562	2	050458	050458 mycobacteri
33	222	10.7	272	5	Q9VJ88	Q9VJ88 drosophila
34	222	10.7	767	2	053435	053435 mycobacteri
35	220.5	10.6	396	10	063450	063450 arabidopsis
36	220.5	10.6	484	2	053394	053394 mycobacteri
37	220	10.6	532	2	006215	006215 mycobacteri
38	219.5	10.6	741	2	006808	006808 mycobacteri
39	218.5	10.5	390	10	Q9M3Y2	Q9M3Y2 tritium ae
40	218.5	10.5	525	2	033284	033284 mycobacteri
41	218.5	10.5	639	2	053952	053952 mycobacteri
42	217.5	10.5	210	10	Q41187	Q41187 arabidopsis
43	217.5	10.5	220	10	Q39367	Q39367 brassica ol
44	213.5	10.3	584	2	053575	053575 mycobacteri
45	212.5	10.2	370	2	053159	053159 mycobacteri

## ALIGNMENTS

RESULT	ID	Q9FCY8	PRELIMINARY;	PRT;	382 AA.
Q9FCY8	Q9FCY8	Q9FCY8	Q9FCY8	Q9FCY8	Q9FCY8
AC	Q9FCY8	Q9FCY8	Q9FCY8	Q9FCY8	Q9FCY8
DT	01-MAR-2001	(TREMblrel. 16, created)			
DT	01-MAR-2001	(TREMblrel. 16, last sequence update)			
DT	01-MAR-2001	(TREMblrel. 16, last annotation update)			
DE	HRPN.				
GN	HRPN.				
OS	Erwinia stewartii.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
CC	Pantoea.				
NC	NCBI_TaxID=66271;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=SS104;				
RA	Frederick R.D., Ahmad M., Majerczak D.R., Arroyo-Rodriguez A.S.,				
RT	Manulis S., Coplin D.L.;				
RT	"Genetic organization of the Pantoea stewartii subsp. stewartii hnp				
RT	gene cluster and sequence analysis of the hrpa, hrpC, hrpN and wtsE				
RT	operons.";				
RL	Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.				
DR	EMBL: AF282857; AAC01466.1; -				
SQ	SEQUENCE 382 AA: 37878 MW: 88A8C723F2DA38BB CRC64;				

Query Match	56.8%; Score 1181.5; DB 2; Length 382;
Best Local Similarity	62.1%; Pred. No. 4.7e-77;
Matches	251; Conservative 44; Mismatches 86; Indels 23; Gaps 11;
QY	1 MSLMTSGAGASTMQISTIGAGGNNGLTGRONAGLGNALGLG-GGNQNDTVQVLA 59
DB	1 MSLMTSGAGASTMQISTIGAGGNNGLTGRONAGLGNALGLG-GGNQNDTVQVLA 57
QY	60 LFGMMNMSMGGGGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLG 119
DB	58 LFGMMNMSMGGGGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLG 109
QY	120 NTSTSTNSPLDQALGINSTSNDSTSGTDSSTSDSDPQQQLKMFSEIMQSLFDDGDCG 179
DB	110 -TTGAGSS-----LGLDPTQTGDDSLSGAGQTSNGMS-PMEQLMKTFADITQSLFDD-QDC 161





QY	67	MSNMGGGGLNG-----GGLGGGGLGN-----GLGSSGGLBGLSNALNDML-----	106
Db	188	FCGNGGAGGIGVAGINGCLOGAGGDDGAGMGCFPGNGNGGAGGAGAAVYAVPGLATPT	247
QY	107	-----GGLMTLTSKSG--GNNTTSTTNSPLDQALGINSTSQNDSTSG----TDSTDS	154
Db	248	PAANGGNGLLIVGYPGATGGAGDANGSALGQAGGAGGAGNASTGGGIGIAQTGTGAGA	307
QY	155	SDPMQLLKMFESEIMSLFDDGDGTGSSSGGKOTPEEQNAYKKGVYDALSLGMLNGL	214
Db	308	G-----GAGDGAAPGNGNGSGVEHT-----GATGS--SAGSNGA	342
QY	215	SQLLNGGLG-----GGGAGNAGTGLDSSSLGKGLQMLSGPV	252
Db	343	TG--GNGGVAPGAGGAGNGGSHVSGSVNTAGAGCKGKGNGGTGAGGP--GGHGGSVLSGPV	399
QY	253	DYQGLGNAVGTGIGMKAGCIGDALNDICGTHRISSTRS----FYVKGDPAAMAKETGEFMDYTP	308
Db	400	GDSGNGGAGGGG--GAGVSATDITAGTGGGNGGNGHGLMTGNGGGGAGGVG-----	449
QY	309	EVEFGPKYQKPGEGVEVTDTRKSNAKALSKRPDDGMTPASMEQFNKAKGIRKPPM--AG	364
Db	450	-----GVG-----GAGAAGALICGGDGG-----GSVNPPIGSGSEAG	480
QY	365	DTGNGNTQARAGAGGSSSLGIDAPMAAGDAIINNALGKICAA	403
Db	481	DGKGGLGAGDGGGRCITFG--QFGAGGAGGAGVGGAGGA	517
RESULT	7		
ID	050415	PRELIMINARY; PRT: 731 AA.	
AC	050415;		
DT	01-JUN-1998 (TREMBLrel. 06, Created)		
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)		
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)		
DE	PKRS-FAMILY PROTEIN.		
GN	RV3388 OR KTV004.46.		
OS	Mycobacterium tuberculosis.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.		
OX	NCBI_TaxId=1773;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-H37RV;		
RX	MEDLINE=98295987; PubMed=9634230;		
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.D.,		
RA	Gordon S.V., Eiglmeier K., Gass S., Barry C.E. III, Tekala F.,		
RA	Badcock K., Basham D., Brown D., Chillingworth J., Connor R.,		
RA	Havens R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,		
RA	Hornby T., Jagels K., Krogsh A., McLean J., Moutle S., Murphy L.,		
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,		
RA	Rutler S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,		
RA	Taylor K., Whitehead S., Barrall B.G.;		
RT	"Deciphering the biology of Mycobacterium tuberculosis from the		
RT	complete genome sequence."		
RL	Nature 393:537-544(1998).		
DR	EMBL: AI009198, CAA15773.1; -.		
DR	HSSP: P03622; Z1FO.		
DR	tubercuList: RV3388; -.		
DR	InterPro: IPR000084; -.		
DR	InterPro: IPR002173; -.		
DR	Pfam: PF00934; PE: 1.		
DR	ProDom: PD001223; -; 1.		
DR	ProSite: PS00583; PFKE_KINASES_1; UNKNOWN.2.		
SO	SEQUENCE 731 AA; 60454 MW; 5A8F07012357D68 CRC64;		

Query Match	12.0%;	Score 249;	DB 2;	Length 731;
Best Local Similarity	28.7%;	Pred. No. 5.6e-10;		
Matches 96;	Conservative 23;	Mismatches 101;	Indels 114;	Gaps 15;
0Y	2	SLNTSGCLASTPQDLSIGGAGGNNGLGTSRKQ-----NAGLGGN-SALGIG	45	

```

Db      402 SAGTGVAS-----GGTCGACGLTGAGHGAGAGAGNOTGTGNGNCGAGNGAGAG 455
QY      46 G--GNODTVNQLAGLLTYGMNNMNMNMGGGGLMGGLG-GGLG-----NGIGSGG 93
Db      456 GOLYNGGDGCGNGAG-----GANINGAGNSDGAAGHAGAGSARLIGAGHGCGDG 508
QY      94 LGEGLSNALNDMLGSLWTLSKGN-----NTSTNSPLD 130
Db      509 AGGNTRAGRAAIACT----GDDGNGNGNGILLSCNAGAGHGAGGSSTATTYTGTPT 564
QY      131 QALGINSTRONDSDSTGTDSTSDSPMQLLIKMFSEIMOSLFEPGDGDTGSSSGKP 190
Db      555 CATGNGNGNGCAGTAGTGTGSG-----GIGNGCAGAGTGCNAG-----602
QY      191 TEGEONAYKKGYTDAL-----SGLMGNCLSQLGNGLG--GGGNGAGTGLDSSLGK 243
Db      603 -----VALSVSTGTLGNGGSGGGLGGGGSLFNGCAGAGVATGTCGNGSGIGPVSQN 657
QY      244 G-----LQNLSEPVDTQQQLNAVGTGIKMGAG 270
Db      658 GKKGCVGAAGLAG---QIQNGSGSGSGGAG 686

RESULT      8
ID      053552 PRELIMINARY; PRT: 1381 AA.
AC      053552;
DT      01-JUN-1998 (TREMBLrel. 06, Created)
DT      01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT      01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE      PGMS-FAMILY PROTEIN.
GN      RV3507 OR MTW023.14.
OS      Mycobacterium tuberculosis.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC      Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX      NCBI_TaxID=11773;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-H37RV;
RX      MEDLINE=98295987; PubMed=9634230;
RA      Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA      Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA      Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA      Davies R.K., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA      Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA      Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA      Rutter S., Seeger K., Skelton S., Squares S., Sulston J.E.,
RA      Taylor K., Whitehead S., Barrett B.G.;
RT      "Deciphering the biology of Mycobacterium tuberculosis from the
RT      complete genome sequence."
RL      Nature 393:537-544(1998).
DR      EMBL: AL022022; CAAL17744.1; -
DR      HSSP: P00778; 2UTL.
DR      TuberculinSc; RV3507; -
DR      InterPro: IPR000084; -
DR      InterPro: IPR002173; -
DR      Pfam: PF00934; PE_1.
DR      Prodom: PD001223; -.
DR      PROSITE: PS00583; PKRB_KINASES.1; UNKNOWN_2.
SQ      SEQUENCE 1381 AA; 110624 MW; CA09676BD07F6482 CRC64;
```

```
Query Match Similarity    11.9%; Score 248; DB 2; Length 1381;  
Best Local Similarity   26.8%; Pred.No. 1.5e-09;  
Matches 117; Conservative 24; Mismatches 180; Indels 116; Gaps 16.  
  
QY      7 GGAATMCAISIGAGGANNGLCTSRONALAGNSAAGLGGQNNDTVNOLAGLTGMNNMM 66  
        I:-I- -::| | | | : | | | | | |  
Db       822 GGAGGTTTTCVNGNCGNAC-DGCGNGCNACGAGSGCDRGGMT----- 864  
  
QY     67 MSMMGGGGIMGGCLGAGCGLANGLGSSGLESLNALNDMLASGLNTTLAGSKGGINNTSTTN 126
```



```

Db 865 -SCASSGSGNGNAGTACSGCAGTGTGSLGCGNGCGNGN-----GGDCGNGAHGTVC 919
QY 127 SPLDQALGINSTSONDSTSGTSDTSDSDP--MOQLIKFSEIMQSLFPGDQDGTGSS 184
Db 920 AQFPAVATSLPTPNCAGANGNGTSGNGAPGAPGPTTGGNAGSOGIGCGGNGGGGK 979
QY 185 SGGR-----OPTBEQNAKKGVTDALSLGMLNGSLQLLNGGGLG-----224
Db 980 GGDADAVNVVFMPTPEQ-----AATGTAGSAGDPTGCGGCTPGSPVAVAPPP 1029
QY 225 -----GGGCGNAGTGL-----DSSLGCGKGLN-----LSGPVDYQQLGNVGTGIM 267
Db 1030 TPTTQVQGGDGGGAGGTGTANDGTATGKGGGCVSILGPP-----GNGGTG---1080
QY 268 KAGIQALNDIGTTHRSSTRFVFNKDRAMAKEIGOFMDQYPEVEGPKQYKPGCEVKT 327
Db 1081 -----GNASATGTNGVANAGNGKGGDGGF-----GAGCGNGAGGSV--T 1119
QY 328 DKSMAKALSKPDDGKMTFASMEQFNKAKMTKRPAGDPTGNGTQARAGSSILGIDAM 387
Db 1120 DGSAGSTAGNGNG-----NATNGTI---AGQPAGNGSAGKGGDGNIAAGA 1166
QY 388 AGDAIN--NNALGKIGA 402
Db 1167 TGTGNGNGNGNGMDGA 1183

RESULT 9
ID 050396 PRELIMINARY; PRT: 588 AA.
AC 050396;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DE 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE PGFS-FAMILY PROTEIN.
GN RV3367 OR MTY004.25.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RL EMBL; AL009198; CAA15752.1;
DR Tuberculist; RV3367;
DR InterPro; IPR002173;
DR Pfam; PF00934; PE; 1.
DR PRODOM; PD001223; -.
DR PROSITE; PS00583; PKB_KINASES_1; UNKNOWN_1.
SQ SEQUENCE 588 AA; 49708 MW; 067B84097F61DAF1 CRC64;

```

```

Query Match 11.9%; Score 247; DB 2; Length 588;
Best Local Similarity 30.4%; Pred. No. 6e-10;
Matches 117; Conservative 16; Mismatches 154; Indels 98; Gaps 21;

```

```

QY 18 GGAGG--NNGGLGTSRONAGIGNSALGIGGNGNDTVNQLAGLLGTGMMMMNMGGGL 75
Db 283 GGAGGDANGNPANTSIANAGAGNGAAG-GDGGANGGAGGAGQAASAGSVGGDGGNGG 341

```

```

QY 76 MGG-----GLGGLGNGLGGSGGLGEGLSNALMDLGGSLNTLGSKGNNTSTNSPLDQ 131
Db 342 AGGTGTGNHAGAG-GAGGAGGARGMTLVG--NGNGGN-----GAAGGGAIGTGC---390
QY 132 ALGINSTSONDSTSGTSDTSDSDPMQQLKMFSEIMQSLFPGDQDGTGSSSGKOPT 191
Db 391 AGGV-PANQGNNSALGT-----QPVGGDGGGCKNGTGG---T 424
QY 192 EGEONAYKKGYTDALSG-LMGNGLSQLLNGGLG--GGGNGAGTGLDSSILGKGLQNL 248
Db 425 GGRGDDGSGGAGGASGLMLNGNGNG--GNGGTGSGGVGNGGIGDAG--CGNATSTS 481
QY 249 SGPDYQQLGNAVGTGIMKAKIQALNDIGTTHRSSTRFVFNKDRAMAKEIGOFMDQY 308
Db 482 SIFPDANG-GNG--GAGGDAG-----HGTGGDGGDGGHAGTGGRGGL-----522
QY 309 EYFGKPYQKPGGQEVKTDKSMAKALSKPDDGKMTFASMEQFNKAKMTKRPAGDPTGN 368
Db 523 -----AGQHANSNGGGGGTGGAGGTG--TPSSGN-----AGGTGT 557
QY 369 GNTQARGAGSSILGIDAMAGDAIN 393
Db 558 GNAIDSTNGCPGSDG---LGDAFN 578

RESULT 10
ID 053557 PRELIMINARY; PRT: 1079 AA.
AC 053557;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE PGFS-FAMILY PROTEIN (FRAGMENT).
GN RV3512 OR MTY023.19.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RL EMBL; AL022022; CAA17749.1;
DR Tuberculist; RV3512;
DR InterPro; IPR002202;
DR PROSITE; PS00318; HMG_COA_REDUCTASE_2; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 1079 AA; 81163 MW; A79718CDB74B97D CRC64;

```

```

Query Match 11.9%; Score 246.5; DB 2; Length 1079;
Best Local Similarity 26.2%; Pred. No. 1.4e-09;
Matches 118; Conservative 26; Mismatches 164; Indels 143; Gaps 16;

```

```

QY 5 TSGLGASTMQISTIGAGGNNGLGTSRONAGIGNSALGIGGNGNDTVNQLAGLLTGM 64
Db 636 TGGAGGNGCGGANGGAGGAGSGGTGGNGAGGADGAGNGNGTGNNGNG-----689
QY 65 MMSNMNMGGL--MGGGLGGGLGNGLGGSGGLGEGLSNALNDMLGGLMT-----LGSKGN 119
Db 690 -----GNGGIAGMGNGAGGTGSGNGNGSG--GNGGNMGNGNSGTGSGDGGAGGN 740

```



```

QY 120 NTSTNSPLDQALGINST-----SONDSTGCTSTSDSPMOQLMFSEIM 169
Db 741 GGAAGTGTGGDGG-GLTGTGTGGTGGTGGGAGGAGADNTAN-----MT 785
QY 170 QSLFGDQDGTGSSG-----GKOPTGEONATKKG-----VTDLASG 208
Db 786 AAGAGDGGNGGDDGGFGGAGAGGGLTGANCTGGCAGAGDGGNGAIGCHPLTDDPGG 845
QY 209 LMGSLQLGNGSLG-----GGGNGMGTGTDGSSLGKGLQNLSCP 251
Db 846 NGGTG-----GNGTGTGTGAGAGISLGGCTGGDGGNGNGTGGEGEVGAG-----893
QY 252 VDYQQLGNAVGTGIGMAGIQAIALNDIGTHRSSTRSEFVNKGRAMAKEIGQFMQYEVF 311
Db 894 -----GTGGAAGNGGDDGTGGTGGDAGAGGTGTGTGGLD-----PRVG 935
QY 312 GKPYQYGPQGEVYTDKSNAAKLSKPDGDMTPASMEQFNKAKMKIRPMAGDTGNGNL 371
Db 936 G-----SGDGTGTGGSGAGAGNGGNG-----GNAGAGGNGNG 967
QY 372 QARGAGSSIGIDAMMAGDAINNMALGKIGA 402
Db 968 GTGAGG--IGGTGNGGDAEPGVPPAGGA 996

RESULT 11
OS3395 PRELIMINARY: PRT; 1538 AA.
AC 053395;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE PGRS-FAMILY PROTEIN.
PE PGRS OR RV3345C OR MTW004.01C-MTW016.45C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.
RL Nature 393:537-544(1998).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN-H37RV;
RA Parkhill J.;
RA Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RA EMBL: AL021841; CAI17117.1;
RA HSSP: P00441; ISOS.
DR Tuberculin; RV3345C;
DR InterPro: IPR000084;
DR InterPro: IPR002173;
DR Pfam: PF00934; PE: 1.
DR ProDom: PD001223;
DR PROSITE: PS00583; PFKB_KINASES_1; UNKNOWN_3.
DR PROSITE: PS01287; RTC; UNKNOWN_1.
SQ SEQUENCE 1538 AA; 129386 MW; 788F0B2095587592 CRC64;

```

```

QY 7 GLASTWQISIGAGGNGNGLTGTSTRONAGLGG-----NSALGLGGGNDPTVN-----54
Db 249 GTG-----GIGGNGGAGLFG-----NGAGAGAGALPCAGALNGSDGNGGTGCGN 298
QY 55 -QLAGLLTGMAMMNMNMGGGGLMG-----GGLAGGLGN-GLAGSGG 93
Db 299 GGRGGLLVG-----NGAGAGAGAGVGGDGGKGGAGDPSPFAVNNAGAGNGHGNGPVGAGG 354
QY 94 LGEISNANLMDLGSINTLGSKGGNNTTSTNPLDQALGINSTSONDSTGTDSTSD 153
Db 355 AG-GLLAGAMACAAATPTSGNGGDDGIGATVANSPL-QAGAGAGNG-GHGLVNGGTGG 411
QY 154 SSDPMQQLKMFSEIMQSLFGDGTGSSSSGKOPTGEONATKKGVTDALSLGNG 213
Db 412 AG-----GAGHAGSTGATGTLAOTPGNGT-----NGAGAGGNGNGNG 450
QY 214 LSQLLGNGLG-----GGGNGMAGTGLDSSLAGKGLQNLGSPVYQQLGNAVGTGIGMK 268
Db 451 GAQ-HGDOGVGKGGAGSGGAGNGGFDATLGSFADGGK-----GNGGKGGDGK 502
QY 269 AGTALNDIGTHRSSTRSEFVNKGRAMAKEIGQFMQYEVFGKPYQYGPQGEVYTD 328
Db 503 AG-----DGGAGAGDVTYLAVNOGAG-----DGGNGGEVGVG 536
QY 329 KSWAKLS-----KPDGDMTPASMEQFNKAKMKIRPMAGDTGN-----GN 370
Db 537 KGGAGVSAHPALNGSAGANCTAPTSGGNGGAGATPTVAGENGAGNGHGSGVGN 596
QY 371 LQARGAGSSL--GIDAMMAGDAINNMALGKIGA 403
Db 597 GGAGAGAGNGVAGTGL-ALNGNGNGNGIGGNGSA 631

RESULT 12
OS33215 PRELIMINARY: PRT; 1660 AA.
AC 0533215;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE PGRS-FAMILY.
PE RV2490C OR MTW008.46C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.
RL Nature 393:537-544(1998).
RN [2]
RN EMBL: AL021246; CAI16067.1;
RN Tuberculin; RV2490C;
RN InterPro: IPR000084;
RN InterPro: IPR002128;
RN InterPro: IPR002173;
RN Pfam: PF00934; PE: 1.
RN ProDom: PD001223;
RN PROSITE: PS00583; PFKB_KINASES_1; UNKNOWN_1.
RN PROSITE: PS01287; RTC; UNKNOWN_1.
SQ SEQUENCE 1660 AA; 133124 MW; 3A89CE12C0FA945 CRC64;

```

Query Match 11.88; Score 246; DB 2; Length 1538;  
 Best Local Similarity 27.28; Pred. No. 2.3e-09;  
 Matches 124; Conservative 32; Mismatches 168; Indels 132; Gaps 21;



Db 521 GGGASVCTGGSGAGDGFVGA----GGNGNAGTGVGVNANGCN-----GCSA 568  
 QY 61 TGMAMNMGCGGGLMGGLGGLGNGLGGSGGGLGEBLSNALN--DMLGSLNT-LGSKG 117  
 Db 569 TG---ALAAVGGAGAGGADATSGTG-GFGAGGSARGLIRLGGAGAGAGASTGVGPG 624  
 QY 118 GNNSTSTNSPLDQALGINST-SQNDSTSGTSTSDSPMQOLLMFSEIMOSLFGDG 176  
 Db 625 GPGGTGASAPFGIAIAGAGAGAGAGTSGATGAGAGDGVFEGT-----AVLGLRG-G 678  
 QY 177 QDGTGSSSGCKPTBEBQNAKKGYTDALSGL--MGNGLSQLG----NGCLGGGGGN 230  
 Db 679 AAGAGGATG-----DGATGAGAGGAGAGAGAGIANFLGFSVLMGAGAGAGTA 725  
 QY 231 AGTGLDSSSLGKGLQNLGSPVDYQQLGNVGTGIGMKAG-----IQALNDIGTHRSST 285  
 Db 726 TGTGNGGAGAGGGG---LSSPV-----ILGIGTGGAGDGGALCYLGGMG----- 769  
 QY 286 RSFVNKGDRAKAREIGQFMDQYEVFGKPOYOKGPGQEVKTDKSMKALSKPDDGMP 345  
 Db 770 ----DGGDGEAVAVGIAVG-----GAG-----GAG----- 788  
 QY 346 ASMEQFNKAGMIRKPAAGTGNGNLQARGAGSSLCITDAMAGDAINNALGKLG 401  
 Db 789 -----GAGG-----AAPTNG-----GAGGN-----GGDALGLVGVGGNG 818

## RESULT 15

053556 PRELIMINARY; PRT; 714 AA.  
 ID 053556  
 AC 053556;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE PGSR-FAMILY PROTEIN.  
 GN RV3511 OR MTW023.18.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 RN NCBI\_TaxID-1773;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE-98295987; PubMed-9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,  
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,  
 Hornsby T., Jagels K., Kirogh A., McLean J., Moule S., Murphy L.,  
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
 Taylor K., Whitehead S., Barrell B.G.;  
 RA \*Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.\*  
 RL Nature 393:537-544(1998).  
 EMBL: AL022022; CAI17748.1;  
 DR Tuberculist; RV3511;  
 DR InterPro: IPR000084;  
 DR InterPro: IPR002173;  
 DR Pfam: PF00934; PE; 1.  
 DR ProDom; PD001223; -; 1.  
 DR PROSITE; PS00583; PFKB\_KINASES.1; UNKNOWN.1.  
 SO SEQUENCE 714 AA; 59990 MW; E66D2A7AF5CF0E1 CRC64;

Query Match 11.4%; Score 236.5; DB 2; Length 714;

Best Local Similarity 28.0%; Pred. No. 4.3e-09;

Matches 120; Conservative 27; Mismatches 161; Indels 121; Gaps 22;

QY 5 TSGAGASTMGISGAGGNGG-LIGTSRONAGTGCNSALGCGGNONNDYVQNLGCLTGM 63  
 Db 215 TGGIG-----GGPGNGGWLIG-----NGRHGG--AGGIGGSGG-----AGNGCGW 254

QY 64 MAMNMGCGGGLMGGLGGLG-----GLGSGGLGEG-----LSNALND 104  
 Db 255 LLGNGIGGAGGTGGAGGTGGMAMLLGCGGTGAGGIGGNGGNGGNGGWLIGNGNG 314  
 QY 105 MLGSLN--TLGSKGNNNTSTNSP---LDQALGINSTSQNDSTSGTSTSDSPMQ 159  
 Db 315 GLGDDGCGTGGHGGNG-----GNPGWLLCTAGCGGNGGAGAGTGTAGGSGGTG----- 364  
 QY 160 QLLMFSEIMOSLFGDQDG-----TGSSSGCKPTBEBQNA--YKGYVDALSGLMNG 213  
 Db 365 -----GDGTGGRGGLLMGAGAGGHTGAGAGAGVNGGAGAGAGAGGNG 410  
 QY 214 -----LSQLGNGGLG-----GGGNGNAGTGLDG--SSLGKGLQNLGSPVDYQQLGNV 261  
 Db 411 GAGGQALLFRGGTGGAGGTGGDGGGGGDFDGTMAIGSTGSGGTGD-----GAP 465  
 QY 262 GTGIGMAGIALNDIGTHRSSTRSFVNKGDRAKAREIGQFMDQYEVFGKPOYOKPG 321  
 Db 466 GNGAGGAG--QLSHSGVAGASGKAGAGGTGNGGAGAGADAPAGSGAMSGTGFAGAG 524  
 QY 322 QEVKTDKSMKALSKPDDGMPASMEQFNKAGMIRKPAAGTGNGNLQAR-GAGSS 380  
 Db 525 -----GDG-----NGGSGASQGGNGGNGGNGGTGKGGTGAG 558  
 QY 381 L-GIDAMMA 388  
 Db 559 MNSIDPLIA 567

Search completed: October 25, 2001, 11:47:11  
 Job time: 192 sec